

hexagon:2610.00162v1

Nearly Minimax Variance Estimation Under Rough Random Design

P. M. Aronow

Patrick Lopatto

AUTHORS' NOTE

This paper was written by the authors with the assistance of large language models, which included suggesting arguments, contributing to drafting and revision, and writing code for computational checks. It represents an update to a prior version of this paper (arXiv:2607.13170v3), with significant improvements to the bounds previously reported. We have taken care to supervise this writing to ensure that the proofs meet a minimum standard of readability and that previous literature is appropriately cited. However, the resulting document remains lacking in exposition and overall coherence. We are nonetheless releasing this manuscript on Hexagon to disseminate the result as quickly as possible, because we believe it may be of interest to the statistics community. We welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

P.L. was partially supported by NSF grant DMS-2450004.

ABSTRACT

We determine, up to a power of $\log n$, the minimax risk for constant conditional variance estimation under rough random design. The unknown design density is bounded above and away from zero, with no smoothness assumption, and the conditional error laws may depend on the covariates and have uniformly bounded fourth moments. For an s -Hölder regression function with $s > 1$ in dimension $d > 4s$, the minimax root-mean-square risk lies, for all sufficiently large n , between $c\Psi_n$ and $C\Psi_n(\log n)^\Gamma$, where $\Psi_n = n^{-2(s+1)/(d+4)}e^{-\kappa\sqrt{\log n}}(\log n)^{(s-1)/(d+4)}$ and the constants $\kappa > 0$ and $\Gamma > 0$ are explicit. In particular, the minimax exponent is $2(s+1)/(d+4)$, the minimax risk is smaller than $n^{-2(s+1)/(d+4)}$ by a stretched-exponential factor whose constant κ is identified, and the rate proposed by Robins is not uniformly attainable over this model class. For $0 < s \leq 1$, we show that the exact minimax rate is $n^{-1/2} \vee n^{-4s/(d+4s)}$; for $s > 1$ and $d \leq 4s$, it is $n^{-1/2}$.

Contents

1	Introduction	1
1.1	Background	1
1.2	Main results	3
1.3	Ideas of the proofs	6
1.4	Related work	9
1.5	Organization and conventions	14
2	Pair differences and elementary upper bounds	15
2.1	Parametric upper bounds by residual projection.	15
2.2	A common risk bound for pair statistics.	16
3	A refined upper bound: resolving below the sampling scale	21
3.1	Setting and fixed quantities	21
3.2	Anchored monomials and local fits	23
3.3	A polynomial estimator of a fitted increment	26
3.4	Unbiased polynomials from a fixed sample	29
3.5	Definition of the estimator	31
3.6	The pilot and its risk bound.	33
3.7	One allocation for bias and variance.	35
3.8	Completion of the upper bound.	37
4	Preliminaries for the lower bounds	38
4.1	The ternary law.	38
4.2	One score-to-risk comparison	39
4.3	The parametric lower bound	39
4.4	Smooth weights	40
5	The case $0 < s \leq 1$	41
5.1	Upper bound.	41
5.2	Lower bound.	41
6	The lower bound	44
6.1	The sample, patches and raw states	44
6.2	Interpolation on one patch	47
6.3	Signed local updates	50
6.4	Cancellation for two observations.	53

<i>Contents</i>	iv
6.5 The generator and its local cancellation	58
6.6 Activity and aliases	62
6.7 A positive path of priors	63
6.8 From local updates to risk	67
6.9 Parameters and proof of Theorem 6.1	69
References	72

Introduction

1.1 Background

We consider a problem that arises in the study of higher-order influence function estimators [46] and semiparametric minimax lower bounds [47] for nonlinear functionals. We observe independent copies of (X, Y) , where $X \in [0, 1]^d$ has an unknown density p bounded above and away from zero, and let

$$f(x) = \mathbb{E}(Y \mid X = x), \quad \text{Var}(Y \mid X = x) = V.$$

The regression function belongs to a fixed s -Hölder ball. Let Q_x denote the conditional law of $\varepsilon = Y - f(X)$ given $X = x$. It has mean zero and variance V , and its fourth moment is uniformly bounded, but it may otherwise depend on x . No smoothness is imposed on p . We aim to estimate V , which satisfies

$$V = \mathbb{E}(Y^2) - \int_{[0,1]^d} f(x)^2 p(x) \, dx. \quad (1.1)$$

Under the uniform Hölder-norm and conditional fourth-moment bounds, the empirical mean of Y^2 estimates $\mathbb{E}(Y^2)$ at the root- n rate. Thus the only potentially non-root- n component in (1.1) is the quadratic regression functional.

For $0 < s < 1$ and $d > 4s$, Robins et al. [46, Section 4] give a close-pair construction with convergence rate $n^{-4s/(d+4s)}$, which uses no smoothness of p . They divide $[0, 1]^d$ into small cubes, select one pair of observations in each cube that contains at least two, and use the identity

$$\mathbb{E} \left[\frac{(Y_i - Y_j)^2}{2} \mid X_i = x, X_j = y \right] = V + \frac{(f(x) - f(y))^2}{2}.$$

By this identity, it is natural to average the squared differences over pairs whose covariates are within distance h . The expected number of such pairs is of order $n^2 h^d$. Averaging the corresponding squared differences gives a stochastic fluctuation bounded in root mean square by $C\{n^{-1/2} + (n^2 h^d)^{-1/2}\}$, while s -Hölder smoothness gives bias $O(h^{2s})$. Balancing the bias with the pair term gives $h \asymp n^{-2/(d+4s)}$; since $d > 4s$, the root- n term is smaller, and the stated rate follows. Robins et al. introduced their construction to show that a random design can give a faster rate than an equally spaced design even when

the design density has no smoothness [46, p. 338]. For s -Hölder regression on an equally spaced d -dimensional grid, the corresponding minimax rate for the root-mean-square risk is $n^{-2s/d} \vee n^{-1/2}$. Robins et al. attribute the rate $n^{-2s/d}$, for $s < d/4$, to Wang et al. and Cai et al. [46, p. 338]. In the lower-bound constructions of these papers, alternatives with a constant variance give the lower bound $n^{-2s/d}$, in one dimension [56, Section 6.2] and in several dimensions [8, Section 6].¹

Richardson and Rotnitzky [44, Section 6] record the following question about the differentiable case, attributed to James Robins: when $s > 1$ but $s/d < 1/4$, does there still exist an estimator attaining $n^{-4s/(d+4s)}$? They write that analogy with other nonparametric estimation problems would suggest that the answer is “yes”. As they also note [44, Section 6, footnote 50], the squared-difference estimator of Robins et al. does not attain this rate for $s > 1$, because it does not exploit the differentiability of the regression function. For $s > 1$, its worst-case bias is generally of order h^2 , rather than h^{2s} . This shows only that the estimator fails to exploit the additional smoothness; it does not rule out that another estimator may attain a better rate. Their question concerns rates of convergence; we study its minimax formulation, which requires a uniform bound on the root-mean-square risk over the model class specified below.

For $s > 1$ and $d > 4s$, we determine the minimax root-mean-square risk up to a power of $\log n$. Let

$$\lambda = \frac{2(s+1)}{d+4}. \quad (1.2)$$

For all sufficiently large n , the minimax root-mean-square risk lies between $c\Psi_n$ and $C\Psi_n(\log n)^\Gamma$, where

$$\Psi_n = n^{-\lambda} e^{-\kappa\sqrt{\log n}} (\log n)^{(s-1)/(d+4)}$$

and the constants $\kappa > 0$ and $\Gamma > 0$ are given explicitly in Theorem 1.1. These bounds identify λ as the minimax root-mean-square exponent. The minimax risk is smaller than $n^{-\lambda}$ by the factor $e^{-\kappa\sqrt{\log n}}$, up to powers of $\log n$, with the same constant κ in the upper and lower bounds, and the ratio of the upper and lower bounds is at most $C(\log n)^\Gamma$. Thus the rate $n^{-4s/(d+4s)}$ is not uniformly attainable over the stated model class, which gives a negative answer to the minimax formulation of Robins’s question. The upper bound also improves on the fixed-design minimax rate. For $0 < s \leq 1$, we give a direct proof of the exact minimax rate $n^{-1/2} \vee n^{-4s/(d+4s)}$. For every $s > 0$ and $d \leq 4s$, ordinary least-squares residuals on a fixed partition attain the parametric rate $n^{-1/2}$. Some of

¹The construction in [8, Section 6] uses cone-shaped bumps, which do not belong to C^s when $s > 1$. For $s > 1$, the same lower bound follows by using rescaled C^∞ bumps that equal one at the grid points and are supported in balls of radius less than half the grid spacing. Taking heights $cn^{-s/d}$ with a sufficiently small fixed $c > 0$ gives uniformly bounded C^s norms, while leaving the prescribed values at the design points, and hence the testing calculation, unchanged.

these results appear in earlier versions of this paper, and related results were obtained in concurrent work; we describe both in Section 1.4.

1.2 Main results

Fix $s > 0$ and an integer $d \geq 1$. Fix an open set $\mathcal{U} \subset \mathbb{R}^d$ containing $[0, 1]^d$, and constants

$$0 < p_- < 1 < p_+ < \infty, \quad L_s > 0, \quad 0 < v_- < v_+ < \infty, \quad v_-^2 < C_4 < \infty. \quad (1.3)$$

We adopt the following notation for Hölder norms. Set

$$\ell = \lceil s \rceil - 1, \quad \alpha = s - \ell \in (0, 1]. \quad (1.4)$$

For a scalar function F on an open set $\Omega \subset \mathbb{R}^d$, define

$$[D^\ell F]_{\alpha; \Omega} = \max_{|\gamma|=\ell} \sup_{\substack{u, v \in \Omega \\ u \neq v}} \frac{|\partial^\gamma F(u) - \partial^\gamma F(v)|}{\|u - v\|^\alpha}, \quad (1.5)$$

$$\|F\|_{C^s(\Omega)} = \sum_{|\gamma| \leq \ell} \|\partial^\gamma F\|_{L^\infty(\Omega)} + [D^\ell F]_{\alpha; \Omega}, \quad (1.6)$$

where $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^d$. When $s \notin \mathbb{N}$, this is the usual $C^{\ell, \alpha}$ norm, while integer s is interpreted as $C^{s-1, 1}$.

For each $n \in \mathbb{N}$, one observes independent and identically distributed pairs $(X_i, Y_i)_{i=1}^n$. Their common distribution is indexed by a parameter $\theta = (p, f, V, Q)$ satisfying the following three conditions. Here, p is a probability density on $[0, 1]^d$, $f: [0, 1]^d \rightarrow \mathbb{R}$ is a regression function, $V > 0$, and $Q = (Q_x)_{x \in [0, 1]^d}$ is a measurable probability kernel from $[0, 1]^d$ to $\mathbb{R}$. Define the regression error by $\varepsilon_i = Y_i - f(X_i)$.

(i) Each X_i has Lebesgue density p on $[0, 1]^d$, and p satisfies

$$p_- \leq p(x) \leq p_+ \quad \text{for Lebesgue-a.e. } x \in [0, 1]^d.$$

(ii) The function f has an extension $\tilde{f}$ to $\mathcal{U}$ such that

$$\|\tilde{f}\|_{C^s(\mathcal{U})} \leq L_s,$$

where the norm is defined in (1.6).

(iii) For $p(x) dx$ -almost every $x \in [0, 1]^d$, the conditional distribution of ε_i given $X_i = x$ is Q_x , and

$$\int_{\mathbb{R}} u Q_x(du) = 0, \quad \int_{\mathbb{R}} u^2 Q_x(du) = V, \quad \int_{\mathbb{R}} u^4 Q_x(du) \leq C_4,$$

with

$$v_- \leq V \leq v_+.$$

The conditional variance V is constant in x , but the conditional distribution Q_x may otherwise depend arbitrarily on x .

Since $V^2 \leq C_4$ by conditional Jensen's inequality, the effective variance interval is

$$[v_-, \min\{v_+, \sqrt{C_4}\}].$$

The assumption $v_-^2 < C_4$ ensures that this interval is nondegenerate.

Let

$$\Theta = \Theta(s, d, \mathcal{U}, p_-, p_+, L_s, v_-, v_+, C_4)$$

denote this class of parameters θ . For an estimator $\hat{V}_n$ of V , define

$$\mathcal{R}_n(\hat{V}_n; \Theta) = \sup_{\theta \in \Theta} \mathbb{E}_\theta (\hat{V}_n - V)^2, \quad R_n(\Theta) = \inf_{\hat{V}_n} \mathcal{R}_n(\hat{V}_n; \Theta),$$

where the infimum is over all measurable functions of the observations $(X_i, Y_i)_{i=1}^n$. We write

$$\mathfrak{r}_n = R_n(\Theta)^{1/2} = \inf_{\hat{V}_n} \sup_{\theta \in \Theta} \left\{ \mathbb{E}_\theta (\hat{V}_n - V)^2 \right\}^{1/2}$$

for the minimax root-mean-square risk.

We now state our four main results. Their constants depend only on the fixed model parameters; c_ϵ may also depend on ϵ .

Theorem 1.1. *Let $s > 1$ and let $d > 4s$ be an integer. Let λ be as in (1.2) and $\ell = \lceil s \rceil - 1$ as in (1.4), and put*

$$\begin{aligned} \tau &= \log \frac{\sqrt{p_+} + \sqrt{p_-}}{\sqrt{p_+} - \sqrt{p_-}}, & \kappa &= 4 \sqrt{\frac{2\tau(s-1)(d-4s)}{(d+4)^3}}, \\ q_s &= \binom{d+\ell}{d}, & \Gamma &= \frac{d(q_s - \frac{1}{2})}{d+4}. \end{aligned} \tag{1.7}$$

There exist $c, C > 0$ such that, for all sufficiently large n ,

$$c \Psi_n \leq \mathfrak{r}_n \leq C \Psi_n (\log n)^\Gamma, \quad \Psi_n = n^{-\lambda} e^{-\kappa \sqrt{\log n}} (\log n)^{\frac{s-1}{d+4}}. \tag{1.8}$$

The upper bound in (1.8) is Theorem 3.1, proved in Section 3, and the lower bound is Theorem 6.1, proved in Section 6.

The two bounds in (1.8) differ by the factor $(\log n)^\Gamma$; we do not know the correct power of $\log n$. The estimator of Section 3 is nearly minimax in the sense that, for all sufficiently large n , its root-mean-square risk is within a factor $C(\log n)^\Gamma$ of the minimax risk. The bounds agree in the exponent λ and in the constant κ of the stretched-exponential factor. In particular,

$$\log \mathfrak{r}_n = -\lambda \log n - \kappa \sqrt{\log n} + O(\log \log n).$$

The constants τ and κ depend on the density bounds only through the ratio p_+/p_- . Here q_s is the dimension of the space of polynomials of degree at most ℓ in d variables. In Theorem 1.1, the constants c and C and the threshold on n depend only on $(s, d, p_-, p_+, L_s, v_-, v_+, C_4)$, and not on $\mathcal{U}$.

Neither proof gives an explicit threshold on n , and apart from κ we have made no attempt to optimize the constants. The upper-bound proof requires $\sqrt{\log n}$ eventually to dominate fixed multiples of $\log \log n$, with constants depending on q_s . Our analysis is asymptotic and does not establish the estimator's performance at moderate sample sizes.

These bounds determine the minimax exponent. We state this as a corollary.

Corollary 1.2. *Let $s > 1$ and let $d > 4s$ be an integer. For every $\epsilon > 0$, there exist $c_\epsilon, C > 0$ such that, for all sufficiently large n ,*

$$c_\epsilon n^{-\lambda-\epsilon} \leq \mathfrak{r}_n \leq Cn^{-\lambda}.$$

The minimax root-mean-square exponent satisfies

$$\lim_{n \rightarrow \infty} \frac{-\log \mathfrak{r}_n}{\log n} = \frac{2(s+1)}{d+4}.$$

Proof. Since $e^{-\kappa\sqrt{\log n}}(\log n)^{(s-1)/(d+4)+\Gamma} \rightarrow 0$ as $n \rightarrow \infty$, the upper bound in (1.8) implies that $\mathfrak{r}_n \leq Cn^{-\lambda}$ for all sufficiently large n . For every $\epsilon > 0$, we have $e^{-\kappa\sqrt{\log n}}(\log n)^{(s-1)/(d+4)} \geq n^{-\epsilon}$ for all sufficiently large n , and then the lower bound in (1.8) gives $\mathfrak{r}_n \geq cn^{-\lambda-\epsilon}$ for all sufficiently large n . Taking logarithms in these two bounds, dividing by $\log n$, and letting first $n \rightarrow \infty$ and then $\epsilon \downarrow 0$, we obtain the limit. $\square$

The separation from the exponent in the question recorded in [44, Section 6] is

$$\frac{4s}{d+4s} - \lambda = \frac{2(s-1)(d-4s)}{(d+4s)(d+4)} > 0. \quad (1.9)$$

The lower bound rules out a uniform bound of order $n^{-4s/(d+4s)}$ on the root-mean-square risk, even after multiplication by any fixed power of $\log n$. The exponent also satisfies

$$\lambda - \frac{2s}{d} = \frac{2(d-4s)}{d(d+4)} > 0,$$

which means that the upper bound is faster than the fixed-design minimax rate. In the other direction, the minimax risk is not of the exact order $n^{-\lambda}$, since the upper bound in (1.8) implies that $n^\lambda \mathfrak{r}_n \rightarrow 0$. In particular, the weaker bound $Cn^{-\lambda}$ recorded in Section 2 is not the minimax rate. The bound $Cn^{-\lambda}$ was proved in the third arXiv version of this paper [3] and, in concurrent work, by Dobriban et al. [13, Proposition 3.4]; see Section 1.4.

For $0 < s \leq 1$, the exact minimax rate can be determined without the constructions used for $s > 1$; see Section 5.

Theorem 1.3. *Let $0 < s \leq 1$ and let $d \geq 1$ be an integer. There exists $C > 0$ such that, for every $n \geq 1$,*

$$C^{-1}(n^{-1/2} \vee n^{-4s/(d+4s)}) \leq \mathfrak{r}_n \leq C(n^{-1/2} \vee n^{-4s/(d+4s)}). \quad (1.10)$$

The residual-projection estimator of Lemma 2.2 attains the parametric rate when $d \leq 4s$. Together with the lower bound in Section 4.3, it yields the following theorem.

Theorem 1.4. *Let $s > 1$ and let $d \leq 4s$ be a positive integer. There exists $C > 0$ such that, for every $n \geq 1$,*

$$C^{-1}n^{-1/2} \leq \mathfrak{r}_n \leq Cn^{-1/2}. \quad (1.11)$$

At $s = 2$, the threshold $d \leq 4s$ of Theorem 1.4 has a precedent in [52, Section 3.1]. For regression functions with bounded second derivatives, i.i.d. Gaussian errors and an i.i.d. design with a continuous positive density on a compact support, root- n accuracy is obtained there when $d \leq 8$, with bounds that hold conditionally on the design on events of probability arbitrarily close to one.

1.3 Ideas of the proofs

The common approximation problem. Both bounds are governed by the same problem of polynomial approximation. Let τ be as in (1.7). On $[p_-, p_+]$, the reciprocal $1/z$ is approximated by polynomials of degree m with uniform error of order $e^{-\tau m}$. This classical fact follows from the Chebyshev expansion of the reciprocal [34, equation (2.4)], and the geometric factor $e^{-\tau}$ cannot be improved [26, Corollary 2.2]. Dually, a signed measure on $[p_-, p_+]$ that represents the value at the exterior point zero of every polynomial of degree at most m has total variation at least of order $e^{\tau m}$, by the growth of Chebyshev polynomials outside the interval [45, Section 2.7, equation (2.37)], and the rule at the Chebyshev nodes attains this order (Lemma 6.4). The upper bound uses the first statement and the lower bound uses the second, and this is the source of the common constant κ .

The upper bound. We modify the close-pair construction of Robins et al. For $s > 1$, the bias of half the squared difference of two responses at distance h comes from the regression increment $f(x') - f(x)$, which is of order h . We subtract a prediction of this increment from the difference of the responses, and we use two predictions computed from independent samples, one in each factor of the square. As in double cross-fitting [40, Section 2], the mean of the product then contains the product of the two mean prediction errors, and the fluctuations of the predictions contribute no bias. As in [35, Section 3], their covariances across evaluation points remain in the variance, which we control by averaging over many pairs (Proposition 2.3). If the predictions resolve f at K cells, the bias is of order $h^2 K^{-2(s-1)/d}$, while the pair statistic fluctuates at order $n^{-1} h^{-d/2}$. At the sampling scale $K = n$, balancing these two terms gives the rate $n^{-\lambda}$. The two-scale estimator of Dobriban et al. [13], which projects out the local polynomial trend in cells of diameter of order $n^{-1/d}$ and uses one close pair per retained cell, attains the same rate; see Section 1.4.

To improve on $n^{-\lambda}$, we resolve f at $K = ne^T$ cells with $T > 0$ tending to infinity, so that most cells contain no observation and a local fit cannot be computed in them. As in Park [42, Section 3.2, equations (9)–(14)] and in our companion paper with Kallus [4, Sections 4.1–4.3], we replace the inverse of each local Gram matrix by a polynomial in that matrix. The population fit is then a polynomial in local moments of p and pf . Each monomial in these moments can be estimated without bias by a U-statistic [22], that is, by an average over distinct observations, even when most cells are empty; for cell counts, this is the falling-factorial estimate of powers of cell probabilities [24, equation (21)]. This allows resolution below the sampling scale (Section 3.4). Related estimates of functionals of an inverse Gram or covariance matrix appear in the higher-order corrections of Robins et al., which use U-statistics of increasing order with inverse Gram matrices computed under initial nuisance estimates [46, Theorems 3.14–3.15], and in the polynomial estimates of Kong and Valiant [25, Section 2.2] and of Chen, Liu and Mukherjee [10, Section 3.2], who estimate quadratic forms involving matrix powers by U-statistics.

We express the prediction as a sum of increments of anchored local fits over nested dyadic cells (Section 3.2). Since consecutive fits reproduce the same Taylor polynomial of f , each increment is small. To keep this Taylor cancellation, we clear the determinant of the parent Gram matrix before we approximate the remaining inverse (Section 3.3). After a known preconditioning, each local Gram matrix has its spectrum in $[p_-, p_+]$, and an approximation of degree m gives an increment error of relative order $e^{-\tau m}$, up to a factor polynomial in m (Lemma 3.4). The variance of the unbiased lift grows factorially in the degree when the cells hold few observations, and this limits the usable degree. We therefore lower the degree as the resolution increases, along a quadratic profile in the logarithm of the number of cells per observation (Section 3.5). With this profile, we may take T of order $\sqrt{\log n}$, that is, a resolution gain of $e^{O(\sqrt{\log n})}$ cells per observation, and we gain the factor $e^{-2(s-1)T/(d+4)}$, which is $e^{-\kappa\sqrt{\log n}}$ up to powers of $\log n$ (Section 3.7). Our companion paper [4, Sections 4.1–4.4] and Park [42, Sections 3.2–3.4] also keep the cancellation in projection increments over nested partitions before approximating the inverses of Gram matrices, and they also lower the degree on finer partitions. Here the increments are anchored at a design point, evaluated on close pairs, and normalized by an estimate of the noise coefficient.

The lower bound. We build on the ternary law of Section 4.1, which already appears in the third version of this paper [3]. Its probabilities are quadratic in the regression value. By (6.1), we see that a centered perturbation of f with variance η^2 changes the law of one response exactly as an increase of V by η^2 does. Hiding a change of the variance by randomizing the regression mean is the principle of the lower-bound proof in [52, Section 4.5], which places Gaussian regression values on smooth bumps, and of the fixed-design lower bounds of Wang et al. [56, Section 6.2], which match finitely many Gaussian moments. With the ternary law, only observations that share a local perturbation of the

mean distinguish the two hypotheses. For $0 < s \leq 1$, independent perturbations of size h^s on patches of scale h suffice, as in the one-dimensional random-design construction of Shen et al. [50, Sections 2.2 and 5], whose design density vanishes between the plateaus of the bumps. Here the ternary law makes the cancellation for one observation exact, with a design density bounded away from zero (Section 5). Derivatives of expectations of bounded statistics of the data along the resulting path are represented by a centered score whose second moment, for $nh^d \leq 1$, is at most a constant times h^{4s} times the expected number of pairs of observations in a common patch, of order $n^2 h^d$, and the choice $h \asymp n^{-2/(d+4s)}$ gives the separation $n^{-4s/(d+4s)}$. For $s > 1$, the same balance would give only $n^{-4s/(d+4s)}$, which is smaller than $n^{-\lambda}$ by a power of n , and we must reduce the contribution of these pairs.

The likelihood of the observations in a patch contains the product of the values of the design density at their positions. Robins et al. use priors in which the design density and the regression function vary together [47, Section 3]. Here this correlation has two roles. First, a joint update of the density and of f acts on the observations in a patch through a function of their positions, provided that function has a separated representation (Section 6.2). As in the third version, we choose this function by cardinal interpolation, so that the covariance of the perturbation of f , evaluated at the observed points, is diagonal (Lemma 6.3); here count two uses a separate rule (Section 6.4). Second, the density factor selects the number of observations in a patch. We integrate the updates of the density against a signed measure that represents evaluation at zero, as in the dual problem above. Among counts at most m , an update then acts only on patches with a prescribed number $r \leq m$ of observations; larger counts can give aliases. The leading dependence of exterior evaluation on the cutoff m is $e^{\tau m}$; the full signed update also has factors that depend on the selected count, the degree of the derivative rule and the separated representation (Lemmas 6.4 and 6.5). Signed measures dual to polynomial approximation were used to construct priors with matching moments by Lepski, Nemirovski and Spokoiny [30, Section 4.4], and for the reciprocal by Wu and Yang [58, Appendix A]. The diagonal covariance imitates an increase of the variance at each observation, which we cancel by decreasing V . What remains are the defects of the interpolation, the remainders of the fixed Taylor rule and of the rule for count two, and the aliases at counts beyond the exactness range of the selectors (Sections 6.5 and 6.6).

These updates are signed. We realize them by a path of probability priors: we give polynomial weights to finite histories of updates, which we organize as heaps of pieces in the sense of Viennot [55]. Conditional annihilation (6.44) keeps the law of the design density fixed along the path, and the centering of the reference marks (Lemma 6.8) makes the mean of its total mass equal to one; the method of bounded differences then concentrates this mass near one. The weights are positive by a deletion induction, which adapts the argument for hard-core partition functions credited to Dobrushin and

presented by Scott and Sokal [49, Section 5.1] (Section 6.7). A tilt by the n th power of the total mass of the density cancels the normalization of the likelihood at a fixed sample size. The third version Poissonized the sample instead and reweighted the priors to cancel the random mass in the Poisson likelihood [3, Section 1.2, Step 3]. Removing the last update computes the score of the path on every data observable, and the local cancellation makes this score small (Proposition 6.10). Using the score-to-risk comparison of Lemma 4.1, we then obtain the lower bound (Section 6.8). This comparison is related to the information inequality [29, Chapter 2, Theorem 5.10] and to the comparison of mixtures of Cai and Low [7, Section 2]; we integrate the derivative of the mean of an estimator along the path and bound the two endpoint biases.

The common balance. Both proofs balance the logarithm x of the number of cells, or patches, per observation against the degree m of the approximation, or the largest selected count. Relative to $n^{-\lambda}$, a larger x improves the upper bound and weakens the lower bound by the factor $e^{-2(s-1)x/(d+4)}$, while the approximation problem enters through $e^{\tau m}$. In the upper bound, mx must be at most a multiple of $\log n$, to leading order, to control the factorial variance of the unbiased lifts. In the lower bound, it must be at least a multiple of $\log n$, to leading order, to make patches with more than m observations rare. In each bound, the optimal choice takes x and m of order $\sqrt{\log n}$, and in both bounds it gives the factor $e^{-\kappa\sqrt{\log n}}$ with the same constant κ . The third version already used a balance of this kind in its lower bound, with an unspecified constant. Park [42, Sections 2 and 3.4] and our companion paper [4, Section 1.3] use the same balance in both of their bounds. The constructions below explain these mechanisms where they are used, including the logarithmic corrections.

1.4 Related work

Variance estimation in nonparametric regression. Difference-based estimators of the variance go back to the first-difference estimator that Brown and Levine [5, Section 1] attribute to von Neumann and, later, Rice [43]. Gasser, Sroka and Jennen-Steinmetz [18] estimate the residual variance from local linear pseudo-residuals, Hall, Kay and Titterton [21] construct difference sequences that are asymptotically optimal, and Dette, Munk and Wagner [11] compare quadratic-form estimators and their sensitivity to the regression mean. Munk et al. [39] study the estimation of the noise variance by differences in multivariate regression with fixed covariates, and Brown and Levine [5] and Wang et al. [56] study the estimation of a variance function by differences under fixed design. Residual-based estimators form a second family. Hall and Marron [20] estimate the variance from squared nonparametric residuals, and Fan and Yao [17] estimate a conditional variance function by local linear smoothing of squared residuals. The residual projection of Lemma 2.2 belongs to this family, and variance estimators based on pseudo-residuals that annihilate

local linear trends were studied by Spokoiny [52, Sections 2.1–2.2]. Under random design, Müller, Schick and Wefelmeyer [38] estimate the error variance by a covariate-matched U-statistic of squared differences. Du and Schick [16] give a covariate-matched version of the pseudo-residual estimator of Gasser et al., which attains the asymptotic performance of residual-based estimators under their conditions. The close-pair construction of Robins et al. recalled in Section 1.1 does not use smoothness of the design density.

Shen et al. [50] study the estimation of the variance under random design, when the standardized errors have a common law independent of X . In the multivariate case, they estimate the variance by a ratio of two U-statistics of squared differences weighted by a product kernel, and they prove that the minimax mean squared error is of order $n^{-8\alpha/(4\alpha+d)} \vee n^{-1}$ when every coordinatewise Hölder index lies in $(0, 1]$, where α is the harmonic mean of these indices [50, Section 4.1]. For isotropic smoothness $s \leq 1$, this is the square of the rate in Theorem 1.3, and the estimator of Proposition 2.1 for $0 < s \leq 1$ has the same ratio form, with the indicator of a common dyadic cube in place of the product kernel. Their multivariate upper bound [50, Section 4.1, Proposition 1] assumes Hölder smoothness at most one in each coordinate, so their results do not resolve the higher-smoothness case considered here. The design densities used for the nonparametric component of their lower bound vanish on sets of positive measure, so that construction does not yield a lower bound for the present class. Conversely, the present constructions for the lower bounds at rates slower than $n^{-1/2}$ use the ternary law, under which the error law Q_x varies with x through $f(x)$, so they do not establish the same lower bounds for a model in which the errors have a common law independent of X . The design densities of the present constructions are bounded above and away from zero, as the present class requires.

Quadratic functionals and higher-order influence functions. By (1.1), estimating V amounts to estimating the quadratic functional $\int f^2 p$. Donoho and Nussbaum [15] develop minimax theory for quadratic functionals in Gaussian models, Laurent [28] constructs efficient estimators of integral functionals of a density from linear and quadratic approximation terms, and Giné and Nickl [19] estimate the integrated square of a density by kernel U-statistics. Quadratic regression functionals are studied in [23, 32]. The non-root- n results in [32, Section 3.3] assume that the design density belongs to a Hölder class of smoothness strictly greater than $2(1 + \epsilon)s$, in addition to being bounded above and away from zero. With bounded responses and under these density conditions, Liu et al. state that the constant variance can be estimated, adaptively over s , with mean squared error of order $(n/\sqrt{\log n})^{-8s/(4s+d)}$ when $s < d/4$ [32, Remark 3.5]. Higher-order influence function estimators are developed in [46, 48], and semiparametric minimax lower bounds by the comparison of two mixtures in [47]; van der Vaart [54] reviews the underlying higher-order tangent spaces. Mukherjee, Newey and Robins [37, Section 2.3] construct empirical higher-order estimators that are semiparametric efficient, under their conditions, without

estimating a smooth density. In the model in which the conditional variance may depend on X and the design density has Hölder smoothness $\beta_g > 0$, Robins et al. conjecture that the rate in their equation (1.3) is minimax, up to logarithmic factors, when β_g is too small for their estimators to attain $n^{-4s/(d+4s)}$. For $\mathbb{E}\{\text{Var}(Y | X)\}$, this rate tends to $n^{-2s/d}$, up to a logarithmic factor, as $\beta_g \rightarrow 0$ [46, p. 338].

Functional estimation under rough design. For doubly robust functionals such as the expected conditional covariance, let s be the average Hölder smoothness of the two nuisance functions. Robins et al. conjecture that their higher-order estimators, built from an initial estimate of the density that is uniformly consistent, attain the rate $n^{-2s/d}$ in probability, up to a power of $\log n$, when $s < d/4$ [46, p. 381]. For a counterfactual mean, Liu, Mukherjee and Robins describe an empirical higher-order construction whose bias is of order $n^{-2s/d}$, up to a power of $\log n$, without smoothness of the density [31, Section 4]. McClean et al. [35, Section 4.2, Theorem 1] construct a double cross-fit estimator of the expected conditional covariance, based on undersmoothed local polynomial regression. Under their moment and support conditions, they bound its expected absolute error by a term of order $n^{-2s/d}$, up to logarithmic factors and stated with O_P notation, when $s < d/4$ and the density satisfies only upper and lower bounds. In concurrent work, Park [42] and our companion paper [4] show that the minimax exponent of such functionals is $2s/d$ below $s = d/4$ under the same assumption on the density; we compare these works with ours below. The results of the companion include the quadratic functional $\int f^2 p$ and the average conditional variance $\mathbb{E}\{\text{Var}(Y | X)\}$ of a bounded response whose conditional variance may depend on X , with exponent $2s/d$ below $s = d/4$ [4, Section 3.2]. In the present model, where the conditional variance is constant, the exponent is instead $4s/(d+4s)$ for $0 < s \leq 1$ and λ for $s > 1$, when $d > 4s$. For this model when $s > 1$ and $d > 4s$, with a design density of Hölder smoothness above an explicit threshold, Dobriban et al. show that a clipped higher-order influence function estimator of Robins et al. has mean squared error at most $Cn^{-8s/(d+4s)}$ if, in addition, the unclipped construction satisfies the uniform integrated bounds in their equation (50) on squared conditional bias and conditional variance [13, Theorem 2.1]. In our lower-bound priors for $s > 1$, the design densities are updated patch by patch and can be discontinuous, so these priors are not supported on a fixed Hölder ball. Our lower bounds therefore do not apply to the Hölder-density classes of Dobriban et al., where their conditional bound is faster; they note this limitation for an earlier version of this paper [13, Section 4].

Polynomial approximation and moment matching. Lower bounds for nonsmooth functionals are often built from two priors whose moments agree up to some order, through the duality between moment matching and best polynomial approximation; see Lepski, Nemirovski and Spokoiny [30, Section 4.4], Cai and Low [7] and Wu and Yang [58]. Wu and Yang use best polynomial approximation of the reciprocal on an interval bounded away from zero [58, Appendix A]. The corresponding estimators replace the functional

by a polynomial and estimate its terms without bias, as in [7, Section 4.1] and [24]. In the present lower bound, the signed measures act on the values of the design density, and in the upper bound, polynomial approximation replaces the inverses of local Gram matrices. Related polynomial estimators of functionals of inverse covariance matrices appear in [10, 25].

Concurrent work. Dobriban, Mukherjee, Robins and Wang [13] study the present model with the additional assumption that the design density is Hölder of some positive order. For $s > 1$ and $d > 4s$, they construct a two-scale estimator. It partitions $[0, 1]^d$ into cells of diameter of order $n^{-1/d}$, projects out the local polynomial trend of degree $\lceil s \rceil - 1$ in suitable cells, and averages the squared normalized contrasts of one close pair per cell. They prove that its mean squared error is at most $Cn^{-4(s+1)/(d+4)}$, that is, that its root-mean-square risk is at most $Cn^{-\lambda}$ [13, Section 3.3, Proposition 3.4]. Their proof of this bound uses only the upper and lower bounds on the design density [13, Appendix A.5]. Their paper was posted on 8 September 2026, one day before the third version of this paper, which contained an estimator with the same bound. Theorem 3.1 improves the bound $Cn^{-\lambda}$ by the factor $e^{-\kappa\sqrt{\log n}}$, up to powers of $\log n$, and Theorem 6.1 shows that the improved bound is optimal over Θ up to a power of $\log n$.

Park [42], in a paper posted on 4 October 2026, studies the expected conditional covariance $\mathbb{E}\{\text{Cov}(A, Y \mid X)\}$ for responses bounded by one in absolute value, conditional means $\mathbb{E}(A \mid X)$ and $\mathbb{E}(Y \mid X)$ in Hölder classes of orders α and β , and an unknown design density between known bounds $0 < g_- < 1 < g_+$. The conditional covariance may depend on X . His Theorem 1 states that the minimax mean squared error is of order n^{-1} when $\alpha + \beta \geq d/2$. For $w = (\alpha + \beta)/d < 1/2$, the theorem bounds it below by a constant multiple of $n^{-2w}(\log n)^w e^{-4\sqrt{\tau w(1-2w)\log n}}$ and above by a constant multiple of $n^{-2w}(\log n)^{w+q_\alpha+q_\beta+1/2} e^{-4\sqrt{\tau w(1-2w)\log n}}$, where τ is computed from $g_\pm$ as in (1.7), and q_α and q_β are defined as q_s , with α and β in place of s [42, Section 2, Theorem 1]. When $\alpha = \beta = s$, his root-mean-square exponent below the threshold is $2s/d$, and his coefficient of $\sqrt{\log n}$ is $(2/d)\sqrt{2\tau s(d-4s)}$. For $s > 1$ and $d > 4s$, our exponent λ exceeds $2s/d$ by $2(d-4s)/\{d(d+4)\}$, and our coefficient is κ . The difference reflects the constant conditional variance of the present model, which gives the identity for squared differences recalled in Section 1.1. Park also reports a lower bound for the average conditional variance of a single binary response [42, Section 2]; in that model, the conditional variance may depend on X , and the lower bound does not transfer to the present model. The two works share the threshold $d = 4s$, the constant τ , and the identification of the coefficient of $\sqrt{\log n}$ in the upper and lower bounds.

Both works use polynomial approximation on the density interval. In both, the approximation of the reciprocal controls the bias of the estimator, and in Park's lower bound it controls the separation; our lower bound uses signed rules for the evaluation of polynomials at an exterior point. Park's estimator starts from a double cross-fit average of

products of local polynomial residuals and subtracts bias corrections on nested partitions. The corrections replace the inverses of the population Gram matrices by Chebyshev matrix polynomials, they keep the cancellation of the parent polynomial trends, they are estimated from distinct observations, and their degrees decrease on finer partitions [42, Section 3]. Our estimator uses polynomial estimates of the regression increment inside a pair statistic, and it relies on the constant conditional variance. Park’s main text describes his lower bound through two mixtures with fixed cube probabilities, which give the same distribution to the observations in each cube that contains at most $r + 1$ observations, where $r \geq 1$ is the number of matched moments, and whose targets differ by a signed reciprocal moment [42, Section 2]. Our lower bound selects counts by signed local updates, which we realize by positive priors at a fixed sample size. Park cites the third version and describes its bounds, which have an unspecified constant in the stretched-exponential factor [42, Sections 1 and 8]; the present version identifies this constant as κ in both bounds. Our comparison is based on the main text of arXiv:2610.05006v1. Its proofs are deferred to a supplement that is not part of the arXiv PDF or source files, as checked on 8 October 2026.

In our companion paper with Kallus [4], we study missing-at-random means, treatment effects and expected conditional covariances in fully nonparametric models under rough design. For two nuisance functions with average Hölder smoothness s , we obtain the exponent $2s/d$ below $s = d/4$, and we identify the coefficient of $\sqrt{\log n}$ in the upper and lower bounds [4, Theorem 2.3]. For the expected conditional covariance, the exponent and the coefficient agree with those of Park. The estimator for the main theorem of the companion also approximates the inverses of local Gram matrices by polynomials on the interval that contains their eigenvalues and estimates the polynomial terms without bias, but it estimates the target directly, without regression fits or close pairs of observations. The lower bound of the companion follows the strategy of the third version, with a Poissonized sample, but its priors give the design density a random phase, and the two mixtures are separated through a Fourier coefficient of the reciprocal of the density.

Earlier versions of this paper. Versions 1 and 2 of arXiv:2607.13170 [1, 2], posted on 14 July and 17 August 2026 under the title “On Rates Attainable under Random Design: A Negative Answer to a Problem of Robins”, proved that $\mathfrak{r}_n \geq cn^{-\beta}$ for all $n \geq 1$ when $s > 1$ and $d > 4s$, where $\beta = \{d(3s + 1) + 8s\} / \{(d + 2s)(d + 4)\}$. Since $\beta < 4s/(d + 4s)$, this lower bound already shows that the rate in the question recorded by Richardson and Rotnitzky is not uniformly attainable. Version 2 also established the exact minimax rate of Theorem 1.3 for $0 < s \leq 1$. Version 3 [3], posted on 9 September 2026 under the title “Nearly-minimax variance estimation under rough random design”, proved that $cn^{-\lambda}e^{-C\sqrt{\log n}} \leq \mathfrak{r}_n \leq Cn^{-\lambda}$ for $s > 1$ and $d > 4s$, with an unspecified constant C in the stretched-exponential factor. These bounds identified the exponent λ , and version 3 also proved Theorem 1.4. Its upper bound used an interpolation estimator built from corrected

response differences, and its lower bound used a Poissonized sample and approximate matching of local moments. The present version identifies the constant κ in both bounds, so that the upper bound improves to $C\Psi_n(\log n)^\Gamma$ and the two bounds differ only by the factor $(\log n)^\Gamma$. It gives different proofs of Theorems 1.3 and 1.4, and its lower bound for $s > 1$ uses signed local updates and positive priors at a fixed sample size.

1.5 Organization and conventions

Section 2 proves the parametric upper bounds by residual projection and the upper bounds for low smoothness by pair differences. Its common risk bound for pair statistics is also used in Section 3, which proves the refined upper bound for $s > 1$ and $d > 4s$. Section 4 collects the inputs common to the lower bounds: the ternary law, one score-to-risk comparison, and the smooth weights; it also contains the proof of the parametric lower bound. Section 5 proves Theorem 1.3, introducing the local cancellation and score argument in its elementary form. Section 6 develops this argument into the lower bound in Theorem 1.1, using interpolation, count selection and positive priors built from signed local updates. Corollary 1.2 is proved immediately after its statement.

All logarithms are natural. Constants may increase between occurrences. For a finite signed measure μ , $\|\mu\|$ is the total mass of $|\mu|$, that is, the total variation of μ without a factor $1/2$.

The constants of Theorem 1.1 are recalled in Theorem 3.1. Apart from these and from the model of Section 1.2, each of Sections 2, 3, 5 and 6 fixes its own notation, and several letters, among them h, m, q, M, N, ρ, μ and Γ , are also used there with local meanings, declared where they occur.

Pair differences and elementary upper bounds

Least-squares residuals directly give all of the parametric upper bounds. For low smoothness, we obtain the remaining upper bounds from half the squared difference of two nearby responses. We first record the resulting rates, which include a consequence of the upper-bound construction for higher smoothness.

Proposition 2.1. *Let $s > 0$, $d \geq 1$, and $t = \min\{s, 1\}$. There exists a measurable estimator based on n observations such that, for every $n \geq 1$,*

$$\sup_{\theta \in \Theta} \{\mathbb{E}_\theta(\widehat{V}_n - V)^2\}^{1/2} \leq C \left(n^{-1/2} \vee n^{-2(s+t)/(d+4t)} \right). \quad (2.1)$$

For $s > 1$, this proves the upper bound in Theorem 1.4 and the simpler bound $Cn^{-\lambda}$ when $d > 4s$. For $0 < s \leq 1$, it is the upper bound in Theorem 1.3. We prove every case with $d \leq 4s$, including equality, in the next lemma. The bound for low smoothness is proved below using pair differences. For $s > 1$ and $d > 4s$, Theorem 3.1 implies the bound $Cn^{-\lambda}$ for all sufficiently large n , since its stretched exponential dominates its logarithmic factor. The constant estimator v_- covers the remaining sample sizes.

We write $\mathcal{O} = [0, 1]^d \times \mathbb{R}$ for the observation space, equipped with its Borel σ -field, and μ_θ for the law of one observation under θ .

2.1 Parametric upper bounds by residual projection

When we fit cellwise polynomials, a fixed fraction of the observations remains available as residual degrees of freedom. The approximation error is small even if the design points are irregular or the local fit is rank deficient. Related variance estimators based on normalized local linear pseudo-residuals are studied in [52, Sections 2.1–2.2], for fixed design and conditionally on an i.i.d. random design.

Lemma 2.2. *For every $s > 0$ and $d \geq 1$, there exists a Borel estimator satisfying*

$$\sup_{\theta \in \Theta} \|\widehat{V}_n - V\|_{L^2} \leq C(n^{-1/2} + n^{-2s/d}), \quad n \geq 1. \quad (2.2)$$

In particular it attains $Cn^{-1/2}$ whenever $d \leq 4s$.

Proof. Let $\ell = \lceil s \rceil - 1$ and $q = \binom{d+\ell}{d}$. For $n \geq 2q$, we set $k = \lfloor (n/(2q))^{1/d} \rfloor$ and partition $[0, 1]^d$ into $K = k^d$ cubes, assigning their faces by a fixed convention. Then $K \asymp n$ and $qK \leq n/2$. Let B_X evaluate a fixed basis of the cellwise degree- ℓ polynomials at $X_1, \dots, X_n$, and let P_X be the orthogonal projection onto its column space. We define

$$A_X = I_n - P_X, \quad d_X = \text{tr } A_X \geq n - qK \geq n/2, \quad \tilde{V}_n = \frac{Y^\top A_X Y}{d_X}.$$

The projection is Borel even at changes of rank, since $P_X = \lim_{\epsilon \downarrow 0} B_X(B_X^\top B_X + \epsilon I_{qK})^{-1} B_X^\top$.

Let $f_X = (f(X_i))_{i \leq n}$. On each cube, the Taylor polynomial centered at a fixed corner of that cube approximates f uniformly to within $CK^{-s/d}$. For $\ell = 0$, this is the Hölder bound; otherwise, it follows by integrating the order- ℓ Hölder remainder along a segment in the cube. Since these polynomials give a vector in the range of P_X , we have

$$\|A_X f_X\|^2 \leq CnK^{-2s/d}, \quad \mathbb{E}(\tilde{V}_n | X) - V = \frac{\|A_X f_X\|^2}{d_X} \leq CK^{-2s/d}.$$

Conditional on X , the errors are independent and centered, with variance V and fourth moments at most C_4 . Expanding the quadratic form of a symmetric matrix A , we find

$$\text{Var}(\epsilon^\top A \epsilon | X) = \sum_i A_{ii}^2 \text{Var}(\epsilon_i^2 | X) + 4V^2 \sum_{i < j} A_{ij}^2 \leq C \text{tr}(A^2).$$

We also have $\text{Var}(2f_X^\top A_X \epsilon | X) = 4V\|A_X f_X\|^2$. Using $A_X^2 = A_X$ and $\text{Var}(U + W) \leq 2\text{Var } U + 2\text{Var } W$, we obtain $\text{Var}(\tilde{V}_n | X) \leq C/n$. Then the conditional mean-square error is at most $C(n^{-1} + K^{-4s/d})$, uniformly in the design points. We clip $\tilde{V}_n$ to $[v_-, v_+]$ and average over X to obtain (2.2). For $n < 2q$, we use the constant estimator v_- and increase C . $\square$

2.2 A common risk bound for pair statistics

We partition $[0, 1]^d$ into dyadic cubes of side h and set $k_h(x, x') = h^{-d}$ when x, x' belong to the same cube of this partition, and $k_h(x, x') = 0$ otherwise, in particular when x or x' lies outside $[0, 1]^d$. On the pair set $\{k_h > 0\}$, we denote points by $w = (x, x')$ and let $M_h(dw) = k_h(x, x')p(x)p(x')dx dx'$. Then

$$1 \leq \int dM_h \leq p_+, \quad M_h\{(x, x') : x \in A\} \leq p_+^2 |A| \quad (A \subset [0, 1]^d \text{ Borel}). \quad (2.3)$$

Indeed, writing $p_I = \mathbb{P}(X \in I)$, we see that the total mass is $h^{-d} \sum_I p_I^2$. By Cauchy-Schwarz and $\sum_I p_I = 1$, we obtain the lower bound, and $p_I \leq p_+ h^d$ gives the upper bound. The second assertion follows by restricting the first integration to A .

A vector field $c(w) = (1, c_2(w), c_3(w))^\top$ represents the affine pair residual $y' + c_2(w)y + c_3(w)$. We allow random coefficients to cover predicted increments; the constant field $(1, -1, 0)^\top$ gives the ordinary response difference. For this field, the ratio $\hat{N}/\hat{D}$ of the statistics defined below is the average of half the squared response differences over pairs of observations in a common cube. Müller, Schick and Wefelmeyer [38] estimate the variance by a covariate-matched U-statistic of this kind, and Shen et al. [50, Section 4.1, equation (19)] use the same ratio form with a product kernel.

Independent pilots are useful because the mean of their product contains the product of their mean prediction errors, without a pilot variance term. The fluctuations of the pilots still matter, but the test-function bound below controls them once many pairs are averaged. In the double cross-fitting of Newey and Robins [40, Section 2], two nuisance functions are likewise estimated from independent samples, which removes the additional nonlinear bias that remains under single cross-fitting. McClean et al. [35, Section 3, Lemma 1] bound the resulting bias by products of mean nuisance errors, and in their analysis the covariances between evaluations of the same nuisance estimates at different points remain in the variance. When the observed response at the first design point of the ordered pair is used in place of $f(x)$, the noise coefficient changes, and the denominator estimates exactly that normalization.

Proposition 2.3 (A common pair-risk bound). *Let $c_{\mathcal{A}}, c_{\mathcal{B}}$ be independent, identically distributed jointly Borel random vector fields on the pair set, with first coordinate one, independent of an evaluation sample of $m \geq 2$ observations. Suppose $m \geq c_* n$ for a fixed $c_* > 0$. Put $\bar{c} = \mathbb{E}c_{\mathcal{A}}$ and $\zeta_{\mathcal{S}} = c_{\mathcal{S}} - \bar{c}$ for $\mathcal{S} \in \{\mathcal{A}, \mathcal{B}\}$. Suppose that, uniformly over $\theta \in \Theta$, for deterministic $b, W \geq 0$ and a fixed $C_p < \infty$,*

$$\begin{aligned} \sup_w |\bar{c}(w)| &\leq C_p, & \sup_w \mathbb{E}|\zeta_{\mathcal{S}}(w)|^2 &\leq C_p W, \\ \left\| \int v^\top \zeta_{\mathcal{S}} dM_h \right\|_{L^2} &\leq C_p \sqrt{W/n} \|v\|_{L^2(M_h)} \quad \text{for every Borel } v \in L^2(M_h; \mathbb{R}^3), \\ r_h &= \bar{c}^\top \mathbf{f}, & \mathbf{f}(w) &= (f(x'), f(x), 1)^\top, & \|r_h\|_\infty &\leq b. \end{aligned} \quad (2.4)$$

For $O_i = (X_i, Y_i)$, let $w_{ii'} = (X_i, X_{i'})$, $z_{ii'} = (Y_{i'}, Y_i, 1)^\top$, $P_0 = \text{diag}(1, 1, 0)$, and $(m)_2 = m(m-1)$. Define

$$\begin{aligned} \hat{N} &= \frac{1}{2(m)_2} \sum_{i \neq i'} k_h(X_i, X_{i'}) c_{\mathcal{A}}(w_{ii'})^\top z_{ii'} z_{ii'}^\top c_{\mathcal{B}}(w_{ii'}), \\ \hat{D} &= \frac{1}{2(m)_2} \sum_{i \neq i'} k_h(X_i, X_{i'}) c_{\mathcal{A}}(w_{ii'})^\top P_0 c_{\mathcal{B}}(w_{ii'}), \end{aligned} \quad (2.5)$$

reading a summand as zero when $k_h = 0$. These statistics are square integrable and $\mathbb{E}\hat{D} \geq 1/2$. If a fixed $a > 0$ satisfies $\mathbb{E}_\theta \hat{D} \geq 2a$ for every θ , then, with clipping denoting projection onto the

displayed interval,

$$\hat{V}_n = \text{clip}_{[v_-, v_+]} \left(\frac{\hat{N}}{\max\{\hat{D}, a\}} \right), \quad (2.6)$$

$$\sup_{\theta \in \Theta} \|\hat{V}_n - V\|_{L^2} \leq C \left[b^2 + (n^{-1/2} + n^{-1}h^{-d/2})(1+W) \right].$$

Here C depends only on the model parameters and (c_*, C_p, a) . In particular, $a = 1/4$ always works.

Proof. Fix θ , and write $\mu = \mu_\theta$. Using the assumptions and (2.3), we obtain $\sup_w \mathbb{E}|c_S(w)|^2 \leq C(1+W)$ and, by Tonelli's theorem, $\mathbb{E}\|\zeta_S\|_{L^2(M_h)}^2 \leq CW$. Then the random test functions used below belong to $L^2(M_h)$ almost surely.

For an evaluation pair $o = (x, y), o' = (x', y')$, let $z = (y', y, 1)^\top$. We take either $A = zz^\top - VP_0$ or $A = P_0$, and define $F = c_{\mathcal{A}}^\top A c_B$ and $T_F = [2(m)_2]^{-1} \sum_{i \neq i'} k_h(X_i, X_{i'}) F(O_i, O_{i'})$. The two resulting statistics are $\hat{N} - V\hat{D}$ and $\hat{D}$, respectively. The assumption on the conditional fourth moments and $|f| \leq L_s$ imply that $\mathbb{E}[\|A\|^2 | X, X'] \leq C$. Since the two pilots and the evaluation sample are independent, we obtain

$$B_F := \iint k_h F^2 d\mu^2, \quad \mathbb{E}B_F \leq C(1+W)^2. \quad (2.7)$$

Note that this bound uses only one second moment from each pilot. By finite mass and Cauchy-Schwarz, we also have $\mathbb{E} \iint k_h |F| d\mu^2 < \infty$. Since $\mathbb{E}[zz^\top | X, X'] = \mathbf{f}\mathbf{f}^\top + VP_0$, we have

$$\mathbb{E}[T_F | \mathcal{A}, \mathcal{B}] = L_B := \frac{1}{2} \int c_{\mathcal{A}}^\top B c_B dM_h, \quad B = \mathbf{f}\mathbf{f}^\top \text{ or } P_0. \quad (2.8)$$

Note that this calculation of the mean does not require independence between pairs, since each ordered pair has law μ^2 . Averaging over the pilots, we obtain

$$\mathbb{E}(\hat{N} - V\hat{D}) = \frac{1}{2} \int r_h^2 dM_h \leq Cb^2, \quad (2.9)$$

$$\mathbb{E}\hat{D} = \frac{1}{2} \int \bar{c}^\top P_0 \bar{c} dM_h \geq \frac{1}{2}. \quad (2.10)$$

In the last inequality, we used that $\bar{c}_1 = 1$ and the lower bound in (2.3).

Both B and $\bar{c}$ are uniformly bounded. Expanding the pilots, we find

$$L_B - \mathbb{E}L_B = \frac{1}{2} \int \{\zeta_{\mathcal{A}}^\top B \bar{c} + \bar{c}^\top B \zeta_{\mathcal{B}} + \zeta_{\mathcal{A}}^\top B \zeta_{\mathcal{B}}\} dM_h.$$

By the test-function bound, each linear term has L^2 norm at most $C\sqrt{W/n}$. For the bilinear term, we condition on $\mathcal{B}$ and apply the same bound with the independent test function $B\zeta_{\mathcal{B}}$. Then the squared L^2 norm of the bilinear term is at most $(CW/n)\mathbb{E}\|B\zeta_{\mathcal{B}}\|_{L^2(M_h)}^2 \leq CW^2/n$. Since $\sqrt{W} \leq 1+W$, we conclude that

$$\|L_B - \mathbb{E}L_B\|_{L^2} \leq Cn^{-1/2}(1+W). \quad (2.11)$$

For completeness, we include the order-two variance bound, which is elementary. For a symmetric $H \in L^2(\mu^2)$, let $H_1(o) = \int H(o, o')\mu(do')$ and $\bar{H} = \int H d\mu^2$. We decompose $H = \bar{H} + h_1(o) + h_1(o') + h_2(o, o')$, where $h_1 = H_1 - \bar{H}$ and h_2 has mean zero in either argument. The two orders are orthogonal; within each order, only equal sets of observation indices contribute. Subtracting conditional means is an orthogonal projection. The order-two Hoeffding decomposition [22] then gives

$$\text{Var} \left(\frac{1}{(m)_2} \sum_{i \neq i'} H(O_i, O_{i'}) \right) = \frac{4}{m} \|h_1\|_2^2 + \frac{2}{(m)_2} \|h_2\|_2^2 \leq \frac{4}{m} \int H_1^2 d\mu + \frac{2}{(m)_2} \int H^2 d\mu^2. \quad (2.12)$$

We now fix pilots with $B_F < \infty$, symmetrize $\bar{F}(o, o') = (F(o, o') + F(o', o))/2$, and set $H = k_h \bar{F}/2$. This symmetrization preserves T_F . By Jensen's inequality and $k_h^2 = h^{-d} k_h$, we have $\int H^2 d\mu^2 \leq h^{-d} B_F/4$. A weighted Cauchy-Schwarz inequality gives

$$H_1(o)^2 \leq \frac{1}{4} \left(\int k_h d\mu(o') \right) \left(\int k_h \bar{F}(o, o')^2 d\mu(o') \right), \quad \int H_1^2 d\mu \leq p_+ B_F/4.$$

Then (2.12) implies that

$$\text{Var}(T_F \mid \mathcal{A}, \mathcal{B}) \leq \left(\frac{p_+}{m} + \frac{1}{2(m)_2 h^d} \right) B_F. \quad (2.13)$$

By Cauchy-Schwarz, we also have $|\mathbb{E}[T_F \mid \mathcal{A}, \mathcal{B}]|^2 \leq p_+ B_F/4$, and (2.7) then proves square integrability. Since $(m)_2 \geq m^2/2$ and $m \geq c_* n$, taking expectations yields

$$\|T_F - L_B\|_{L^2} \leq C(n^{-1/2} + n^{-1} h^{-d/2})(1 + W). \quad (2.14)$$

Finally, we note that clipping is 1-Lipschitz and fixes V , and that

$$\frac{\hat{N}}{\max\{\hat{D}, a\}} - V = \frac{\hat{N} - V\hat{D} - V(a - \hat{D})_+}{\max\{\hat{D}, a\}}.$$

Since $\mathbb{E}\hat{D} \geq 2a$, we have $(a - \hat{D})_+ \leq |\hat{D} - \mathbb{E}\hat{D}|$, and it follows that

$$|\hat{V}_n - V| \leq a^{-1} \{|\hat{N} - V\hat{D}| + v_+ |\hat{D} - \mathbb{E}\hat{D}|\}. \quad (2.15)$$

Combining this bound with the two fluctuation bounds and (2.9), we obtain (2.6). The reduction (2.15) is similar to that of Shen et al. [50, Appendix A.1, equation (28)], who control the error of their ratio through the numerator minus the target times the denominator, on an event where the denominator is bounded below. $\square$

Proof of Proposition 2.1 for $0 < s \leq 1$. For $n \geq 2$, we apply Proposition 2.3 with all $m = n$ observations and the deterministic pilots $c_{\mathcal{A}} = c_{\mathcal{B}} = (1, -1, 0)^\top$. Then $W = 0$, and (2.5) becomes

$$\hat{N} = \frac{1}{2(n)_2} \sum_{i \neq i'} k_h(X_i, X_{i'}) (Y_i - Y_{i'})^2, \quad \hat{D} = \frac{1}{(n)_2} \sum_{i \neq i'} k_h(X_i, X_{i'}).$$

The ratio $\hat{N}/\hat{D}$ is a variant of the construction of Robins et al. [46, Section 4], which selects one pair of observations from each cube that contains at least two. Our ratio averages half the squared differences over all pairs in common cubes, as in the ratio estimator of Shen et al. [50]. Here $r_h = f(x') - f(x)$ has norm at most Ch^s . Since $\mathbb{E}\hat{D} = \int dM_h \geq 1$, we may take $a = 1/2$ in (2.6), which gives RMS risk at most $C(h^{2s} + n^{-1/2} + n^{-1}h^{-d/2})$ for the clipped ratio. This statistic is Borel because its kernels are Borel and clipping is continuous. Let h be a dyadic side length such that $n^{-2/(d+4s)}/2 < h \leq n^{-2/(d+4s)}$. Then the risk is at most $C(n^{-1/2} \vee n^{-4s/(d+4s)})$, including $s = 1$ and $d = 4s$. The remaining case $n = 1$ is covered by the estimator v_- , after we increase C . $\square$

A refined upper bound: resolving below the sampling scale

Throughout this section, we assume that $s > 1$ and that $d > 4s$ is an integer. We construct one estimator from two predictions of a regression increment and an independent sample of nearby pairs. By resolving the regression function below the sampling scale, we obtain the following improvement.

Theorem 3.1. *Let $s > 1$ and let $d > 4s$ be an integer. Put $L = \log n$ and*

$$\lambda = \frac{2(s+1)}{d+4}, \quad \tau = \log \frac{\sqrt{p_+} + \sqrt{p_-}}{\sqrt{p_+} - \sqrt{p_-}},$$

$$\kappa = 4\sqrt{\frac{2\tau(s-1)(d-4s)}{(d+4)^3}}, \quad q_s = \binom{d + \lceil s \rceil - 1}{d}.$$

There exist $C < \infty$ and $n_0 \in \mathbb{N}$, depending only on $(s, d, p_-, p_+, L_s, v_-, v_+, C_4)$, such that for every integer $n \geq n_0$

$$\mathfrak{r}_n \leq C n^{-\lambda} e^{-\kappa\sqrt{L}} L^{\frac{s-1}{d+4} + \frac{d(q_s-1/2)}{d+4}}. \quad (3.1)$$

Recalling that $\Psi_n = n^{-\lambda} e^{-\kappa\sqrt{L}} L^{(s-1)/(d+4)}$ and $\Gamma = d(q_s - \frac{1}{2})/(d+4)$, we see that the bound (3.1) reads $\mathfrak{r}_n \leq C \Psi_n L^\Gamma$. The constants do not depend on the extension domain $\mathcal{U}$; the only property of the domain used in bounding the Taylor remainder is the convexity of $[0, 1]^d$. We complete the proof in Section 3.8.

3.1 Setting and fixed quantities

For pairs in one cube of side h , the estimator uses products of two corrected response differences, with corrections given by independent predictions of the increment $f(x') - f(x)$. Resolving the regression function at K cells yields a bias of order $h^2 K^{-2(s-1)/d}$ and a pair error $n^{-1} h^{-d/2}$. When $K = ne^T$, balancing the two improves the rate $n^{-\lambda}$ at the sampling scale by the factor $e^{-2(s-1)T/(d+4)}$. We achieve $T \asymp \sqrt{\log n}$ using unbiased polynomials of local moments, whose degrees are chosen to control their factorial variances. Kong and Valiant [25, Section 2.2] approximate a scalar functional of an unknown covariance matrix by a polynomial in its powers, and they estimate the resulting quadratic forms

without bias by sums over distinct observations. Chen, Liu and Mukherjee [10, Section 3.2] approximate quadratic forms in an inverse covariance matrix by Chebyshev polynomials and estimate the resulting quadratic forms in matrix powers by U-statistics. The higher-order influence function estimators of Robins et al. [46, Theorems 3.14–3.15] use inverse Gram matrices computed under initial nuisance estimates, together with U-statistics of increasing order, to correct the bias of an estimator of a functional.

We write $|\cdot|$ for the Euclidean norm of a vector and $\|\cdot\|$ for the operator norm of a matrix; the matrix blocks of a moment tuple are measured in the Frobenius norm. Let $\ell = \lceil s \rceil - 1$ and

$$q_s = \binom{d+\ell}{d}, \quad r = q_s - 1, \quad L = \log n. \quad (3.2)$$

The constants for the approximation of the inverse and for the allocation are

$$\tau = \log \frac{\sqrt{p_+} + \sqrt{p_-}}{\sqrt{p_+} - \sqrt{p_-}}, \quad \rho = e^{-\tau} \in (0, 1), \quad (3.3)$$

$$\begin{aligned} \vartheta &= (s-1)/d, & \beta &= \vartheta/\tau, \\ \omega &= \frac{d-4s}{d(d+4)}, & \eta &= \frac{2(d+4s)}{d(d+4)}, \end{aligned} \quad (3.4)$$

$$a_* = \sqrt{8\omega/\beta}. \quad (3.5)$$

In particular, we have

$$\lambda = \frac{2(s+1)}{d+4}, \quad \kappa = \frac{2(s-1)}{d+4} a_* = 4 \sqrt{\frac{2\tau(s-1)(d-4s)}{(d+4)^3}}. \quad (3.6)$$

Below, we write C, c for fixed positive constants. They depend only on $\mathfrak{p} = (s, d, p_-, p_+, L_s, v_+, C_4)$, and their values may change between lines. Together with the model moment bounds, the feature constant C_ψ of Lemma 3.2 and the analytic constants M_1, C_1 of Lemma 3.4 determine the variance constant $C_0 \geq 1$, which is fixed sufficiently large in Proposition 3.6.

Model bounds

Let μ be the law of $O = (X, Y)$, and let $\varepsilon = Y - f(X)$. By the assumptions of the model, we have

$$\sup_{[0,1]^d} |f| \leq L_s, \quad (3.7)$$

$$\mathbb{E}[(1 + |Y|)^2 \mid X] \leq M_2 := 2(1 + L_s^2 + v_+). \quad (3.8)$$

For the second bound, note that $\mathbb{E}[Y^2 | X] = f(X)^2 + V$ and $(1 + |Y|)^2 \leq 2(1 + Y^2)$. By Taylor's formula, for all $x, z \in [0, 1]^d$, we have

$$\left| f(z) - \sum_{|\gamma| \leq \ell} \frac{\partial^\gamma f(x)}{\gamma!} (z - x)^\gamma \right| \leq C_{\text{Tay}} |z - x|^s, \quad C_{\text{Tay}} = d^\ell L_s / \ell!. \quad (3.9)$$

To verify the endpoint convention for integer s , we apply the integral form of Taylor's formula in one dimension to $F(t) = \tilde{f}(x + t(z - x))$. The remainder after degree ℓ is

$$\int_0^1 \frac{(1-t)^{\ell-1}}{(\ell-1)!} \{F^{(\ell)}(t) - F^{(\ell)}(0)\} dt.$$

By the Hölder bound and the identity $\sum_{|\gamma|=\ell} \ell! / \gamma! = d^\ell$, its integrand is bounded by $d^\ell L_s |z - x|^s$, and the weight has mass $1/\ell!$. Since the segment stays in the convex cube, the extension domain does not enter these constants. Finally, since $V^2 \leq C_4$, we have

$$V \in I_V := [v_-, \min\{v_+, \sqrt{C_4}\}]. \quad (3.10)$$

Cells and fixed samples

For $j \geq 0$, let $\mathcal{D}_j$ be the set of half-open dyadic cubes of side 2^{-j} in $[0, 1]^d$, let z_I be a cell's corner, and set $K_j = 2^{jd} = |I|^{-1}$. We denote the cell containing x by $I_j(x)$. Any two cells are either disjoint or nested, and

$$p_- K_j^{-1} \leq \mathbb{P}(X \in I) \leq p_+ K_j^{-1} \quad (I \in \mathcal{D}_j). \quad (3.11)$$

We divide the first $3n_b$ observations into independent blocks $\mathcal{A}, \mathcal{B}, \mathcal{P}$, where

$$n_b = \lfloor n/3 \rfloor, \quad \Lambda = n/6. \quad (3.12)$$

The first two blocks are used for the pilots, and the third for the pairs; the remaining observations are discarded. For a block $\mathcal{S}$ of size m , we define the factorial sums

$$\Sigma_k(\phi; \mathcal{S}) = \sum_{\substack{1 \leq i_1, \dots, i_k \leq m \\ i_1, \dots, i_k \text{ distinct}}} \phi(O_{i_1}, \dots, O_{i_k}), \quad (m)_k = m! / (m - k)!. \quad (3.13)$$

The sum is zero when $k > m$, and $\Sigma_0(c; \mathcal{S}) = c$. The admissibility condition below ensures that the denominators are positive.

3.2 Anchored monomials and local fits

We index coordinates by $\mathcal{I} = \{\gamma \in \mathbb{N}_0^d : 1 \leq |\gamma| \leq \ell\}$, taken in a fixed order, and let $q(v) = (v^\gamma)_{\gamma \in \mathcal{I}}$. For $I \in \mathcal{D}_j$ and $x \in I$, we define

$$\psi_{I,x}(z) = q(2^j(z - x)), \quad B_I(x) = \int_{[0,1]^d} q(w - u) q(w - u)^\top dw, \quad u = 2^j(x - z_I). \quad (3.14)$$

We abbreviate $\psi_{j,x} = \psi_{I_j(x),x}$ and $B_j(x) = B_{I_j(x)}(x)$. Note that the matrix $B_I(x)$ is known; its (γ, δ) entry equals

$$\prod_{i=1}^d \frac{(1 - u_i)^{\gamma_i + \delta_i + 1} - (-u_i)^{\gamma_i + \delta_i + 1}}{\gamma_i + \delta_i + 1}.$$

The coordinate transport is the fixed diagonal matrix

$$\mathcal{T}_j = \text{diag}_{\gamma \in \mathcal{I}}(2^{-|\gamma|}) \quad (j \geq 1), \quad \mathcal{T}_0 = 0. \quad (3.15)$$

In part (3) of the next lemma, we normalize the local Gram matrix by its Lebesgue counterpart $B_I(x)$, so that the density bounds place its spectrum in $[p_-, p_+]$. Park [42, Section 3.2, equation (9)] obtains the corresponding spectral interval with a block basis orthonormal for Lebesgue measure. Our companion paper with Kallus [4, Section 4.2, proof of Lemma 4.3] uses a basis orthonormal for normalized Lebesgue measure on each cell and includes the reciprocal cell volume in its Gram matrix.

Lemma 3.2. *There are fixed $c_B, C_B > 0$ and $C_\psi \geq 1$, depending only on (d, s) , such that the following holds at all cells and anchors.*

(1) *The components of $\psi_{I,x}$ form a basis of the degree-at-most- ℓ polynomials vanishing at x , and*

$$K_j \int_I \psi_{I,x} \psi_{I,x}^\top dz = B_I(x), \quad c_B I_r \preceq B_I(x) \preceq C_B I_r. \quad (3.16)$$

(2) *We have $\sqrt{r}, \sqrt{C_B/c_B} \leq C_\psi$, $\sup_I |\psi_{I,x}| \leq C_\psi$,*

$$|\psi_{I,x}(x')| \leq C_\psi 2^j |x' - x| \quad (x' \in I), \quad (3.17)$$

and on I , $|B_I(x)^{-1} \psi_{I,x}|, \|B_I(x)^{-1} \psi_{I,x} \psi_{I,x}^\top\|_F \leq C_\psi^2$.

(3) *If $p_- \leq p \leq p_+$ almost everywhere on I and $H = K_j \int_I p \psi_{I,x} \psi_{I,x}^\top dz$, then*

$$p_- B_I(x) \preceq H \preceq p_+ B_I(x). \quad (3.18)$$

Thus $B_I(x)^{-1} H$ is similar to a symmetric matrix with spectrum in $[p_-, p_+]$, with similarity condition number at most $\chi := \sqrt{C_B/c_B}$.

(4) *On all of $\mathbb{R}^d$, for $j \geq 1$,*

$$\psi_{j-1,x} = \mathcal{T}_j^\top \psi_{j,x}, \quad \|\mathcal{T}_j\| \leq 1/2. \quad (3.19)$$

(5) *The features and known inverse $B_I(x)^{-1}$ are continuous in their arguments on each fixed cell and globally Borel; the transports are constant.*

Proof. The centered monomials form a basis of the space of anchored polynomials, and the Gram identity follows from a change of variables on the cell. If a coefficient vector

is nonzero, then the corresponding polynomial is nonzero, and its squared integral is positive. Since u ranges over the compact cube $[0, 1]^d$, the constants c_B, C_B can be chosen uniformly by continuity, and $B_I(x)^{-1}$ is continuous. Part (2) holds after we increase the single fixed constant C_ψ , since the monomials, their gradients and this inverse are bounded. Part (3) follows from the density bounds: writing $B = B_I(x)$, we see that $S = B^{-1/2}HB^{-1/2}$ has spectrum in $[p_-, p_+]$ and that $B^{-1}H = B^{-1/2}SB^{1/2}$. Scaling each centered monomial gives (4), because $|\gamma| \geq 1$. Finally, (5) follows from the explicit formulas on each cell and the fact that the cells of each level form a finite Borel partition. $\square$

Preconditioned moments and the anchored fit

For a parameter $\theta = (p, f, V, Q) \in \Theta$, we define

$$\begin{aligned} H_j(x) &= K_j \int_{I_j(x)} p(z) \psi_{j,x}(z) \psi_{j,x}(z)^\top dz, & G_j(x) &= B_j(x)^{-1} H_j(x), \\ \beta_j(x) &= B_j(x)^{-1} K_j \int_{I_j(x)} p(z) \psi_{j,x}(z) f(z) dz, & m_j(x) &= B_j(x)^{-1} K_j \int_{I_j(x)} p(z) \psi_{j,x}(z) dz. \end{aligned} \quad (3.20)$$

These quantities are bounded uniformly in the level and the anchor, and they are Borel by dominated convergence on each cell. Using the similarity in Lemma 3.2, we obtain

$$\|G_j\| \leq \chi p_+, \quad \|G_j^{-1}\| \leq \chi p_-^{-1}, \quad p_-^r \leq \det G_j \leq p_+^r. \quad (3.21)$$

For a formal anchor value $y \in \mathbb{R}$, we set

$$b_j(x; y) = \beta_j(x) - y m_j(x) = B_j(x)^{-1} K_j \int_{I_j(x)} p(z) \psi_{j,x}(z) (f(z) - y) dz. \quad (3.22)$$

At the root, the parent values are $G_{-1} = I_r$ and $\beta_{-1} = m_{-1} = 0$. Since $p_- < 1 < p_+$, the spectrum of this parent lies in the same interval. We define the moment tuples by

$$\theta_j(x) = (G_j, G_{j-1}, \beta_j, \beta_{j-1}, m_j, m_{j-1})(x) \in E. \quad (3.23)$$

The fixed finite-dimensional space E carries the Frobenius norm on its matrix coordinates and the Euclidean norm on its vector coordinates. The population fit and the increments are defined by

$$a_j(x) = G_j(x)^{-1} b_j(x; f(x)), \quad a_{-1} = 0, \quad \delta_j(x) = a_j(x) - \mathcal{T}_j a_{j-1}(x). \quad (3.24)$$

We note that the known preconditioner cancels in the fit: $a_j = H_j^{-1} K_j \int_{I_j(x)} p \psi_{j,x} (f - f(x))$.

Lemma 3.3. *Uniformly over parameters, levels and anchors, the following holds. (1) For $x' \in I_j(x)$,*

$$|f(x') - f(x) - \psi_{j,x}(x')^\top a_j(x)| \leq C |x' - x| 2^{-j(s-1)}, \quad (3.25)$$

(2) The increments satisfy

$$|\delta_j(x)| \leq C2^{-js}. \quad (3.26)$$

(3) For every $J \geq 0$ and $x' \in \mathbb{R}^d$,

$$\psi_{J,x}(x')^\top a_J(x) = \sum_{j=0}^J \psi_{j,x}(x')^\top \delta_j(x). \quad (3.27)$$

The constants depend only on (d, s, p_-, p_+, L_s) .

Proof. The nonconstant part of the Taylor polynomial is $P_x = \psi_{j,x}^\top c_j$, where $c_{j,\gamma} = 2^{-j|\gamma|} \partial^\gamma f(x) / \gamma!$. Then $c_j = \mathcal{T}_j c_{j-1}$ for $j \geq 1$. Let $q_x = f - f(x) - P_x$. By the normal equations, we have

$$a_j = c_j + e_j, \quad e_j = H_j^{-1} K_j \int_{I_j(x)} p(z) \psi_{j,x}(z) q_x(z) dz. \quad (3.28)$$

Then (3.9) implies that $|q_x(z)| \leq C|z - x|^s$. Combining this bound with $\|H_j^{-1}\| \leq (p_- c_B)^{-1}$ and the boundedness of the features, we find

$$|e_j| \leq C2^{-js}. \quad (3.29)$$

The error of the fit equals $q_x(x') - \psi_{j,x}(x')^\top e_j$. Since $|x' - x| \leq \sqrt{d}2^{-j}$, the Taylor bound, together with the Lipschitz bound on the features, implies (3.25). For $j \geq 1$, we have $\delta_j = e_j - \mathcal{T}_j e_{j-1}$, which proves (3.26). At the root, the increment bound follows from $|f - f(x)| \leq 2L_s$ and the bound on the inverse Gram matrix. Finally, by (3.19), we have $\psi_{j,x}^\top \delta_j = \psi_{j,x}^\top a_j - \psi_{j-1,x}^\top a_{j-1}$, and the telescoping identity follows by summation. $\square$

3.3 A polynomial estimator of a fitted increment

Recall that the fits at consecutive levels reproduce the same Taylor polynomial. To preserve this cancellation, we clear the determinant of the parent before approximating the remaining inverse. Our companion paper uses the same device for products of two projection residuals: it clears the determinants of the parent Gram matrices and approximates the remaining inverse and reciprocal determinants by one joint series [4, Sections 4.1–4.2]. Park [42, Section 3.3] uses matrix polynomials that annihilate the polynomial trends of the parent exactly, in a multiresolution estimator of the expected conditional covariance. Let $t(z) = (2z - p_- - p_+) / (p_+ - p_-)$ and $\kappa_0 = (p_- p_+)^{-1/2}$. For a free square matrix M , we define the formal series

$$g_\zeta(M) = \kappa_0(1 - \rho^2 \zeta^2) \{I + 2\rho \zeta t(M) + \rho^2 \zeta^2 I\}^{-1},$$

whose denominator has constant term I . Let

$$P_m(A, C) = \sum_{k=0}^m [\zeta^k] \{g_\zeta(A) \det g_\zeta(C)\}. \quad (3.30)$$

Its degree in the matrix entries is at most m , since the recurrence for the inverse produces coefficients whose degree is at most their order in the series, and the determinant and the product preserve this bound.

For integers $m_j \geq 0$, we set $D_j = m_j + q_s$. Writing $G = G_j(x)$ and $G^- = G_{j-1}(x)$, we define

$$\begin{aligned} \mathcal{N}_j(x; y) &= \det(G^-) b_j(x; y) - G \mathcal{T}_j \operatorname{adj}(G^-) b_{j-1}(x; y), \\ F_j(x; y) &= P_{m_j}(G, G^-) \mathcal{N}_j(x; y). \end{aligned} \quad (3.31)$$

Here $\operatorname{adj}(C)$ is the adjugate, which satisfies $\operatorname{adj}(C)C = \det(C)I$, and $\operatorname{adj}(C) = 1$ for $r = 1$. The root convention gives $\mathcal{N}_0 = b_0$ and $F_0 = P_{m_0}(G_0, I)b_0$. We regard the right sides as polynomials $F_{j,x}(\theta; y)$ in the free moment tuple $\theta \in E$. Their coefficients are known and, since $b = \beta - ym$, affine in the formal variable y . For $x' \in I_j(x)$, the polynomial to be lifted is

$$\mathcal{P}_{x,x'}(\theta; y) = \sum_{j=0}^J \psi_{j,x}(x')^\top F_{j,x}(\theta; y), \quad \theta = (\theta_0, \dots, \theta_J). \quad (3.32)$$

We write $D^k \Phi(\theta)[\gamma_1, \dots, \gamma_k]$ for the mixed derivative of $\Phi(\theta + \sum_i t_i \gamma_i)$ at $t = 0$, with the convention $D^0 \Phi = \Phi$.

Lemma 3.4 (The increment polynomial). *There are fixed $C_{\text{inc}}, M_1, C_1 > 0$, depending only on (d, s, p_-, p_+, L_s) , with the following properties at every population tuple, level and anchor, for arbitrary nonnegative degrees m_j .*

(1) For every real y ,

$$G^{-1} \det(G^-)^{-1} \mathcal{N}_j(x; y) = G^{-1} b_j(x; y) - \mathcal{T}_j(G^-)^{-1} b_{j-1}(x; y), \quad (3.33)$$

and hence

$$|\mathcal{N}_j(x; f(x))| \leq \chi p_+^{r+1} |\delta_j(x)| \leq C_{\text{inc}} 2^{-js}. \quad (3.34)$$

(2) The approximation error satisfies:

$$|F_j(x; f(x)) - \delta_j(x)| \leq C_{\text{inc}} 2^{-js} (m_j + 1)^r \rho^{m_j}. \quad (3.35)$$

(3) The polynomial $F_{j,x}(\theta; y)$ has degree at most $D_j = m_j + r + 1$. Every monomial contains one vector-moment coordinate, to first power, and has coefficient constant or linear in y . Thus each observation's response enters its factorial lift at most to first power.

(4) Uniformly for $|y_0| \leq L_s$, $\Phi \in \{F_{j,x}(\cdot; 0), -\partial_y F_{j,x}, F_{j,x}(\cdot; y_0)\}$, $k \geq 0$ and all directions in E ,

$$|D^k \Phi(\theta_j(x))[\gamma_1, \dots, \gamma_k]| \leq M_1 k! C_1^k \prod_i |\gamma_i|. \quad (3.36)$$

The constants M_1, C_1 depend only on $(r, p_-, p_+, C_\psi, L_s)$, not on the level or degrees.

Proof. For $z \in [p_-, p_+]$, let ϕ be such that $t(z) = \cos \phi$. Summing two geometric series, we obtain the Chebyshev identity

$$g_\zeta(z) = \kappa_0 \left\{ 1 + 2 \sum_{k \geq 1} (-\rho \zeta)^k \cos(k\phi) \right\}, \quad |\zeta| < \rho^{-1}.$$

Its coefficients have modulus at most $2\kappa_0 \rho^k$, and a direct substitution of ρ shows that $g_1(z) = z^{-1}$. The identity is the weighted generating function of the Chebyshev polynomials [12, Eq. 18.12.7]; for generating functions, see also [33, Section 5.2.1] and [45, Section 1.5]. At $\zeta = 1$, after an affine change of variables, it is the classical Chebyshev expansion of the reciprocal [33, equation (5.14)], [34, equation (2.4)]. By Lemma 3.2(3), each population Gram matrix is similar to a symmetric matrix with spectrum in this interval, through a similarity with condition number at most χ . The bound on its matrix coefficients then gains only the fixed factor χ . Since the determinant is invariant under similarity, its r scalar factors satisfy the original bounds. Convolving these bounds, we find

$$\|[\zeta^k] \{g_\zeta(A) \det g_\zeta(C)\}\| \leq \chi(2\kappa_0)^{r+1} \binom{k+r}{r} \rho^k.$$

For these population matrices, the series at one equals $A^{-1} \det(C)^{-1}$. Using $\binom{m+1+i+r}{r} \leq (r+1)(m+1)^r \binom{i+r}{r}$ and $\sum_{i \geq 0} \binom{i+r}{r} \rho^i = (1-\rho)^{-r-1}$, we obtain

$$\|P_m(A, C) - A^{-1} \det(C)^{-1}\| \leq C(m+1)^r \rho^m. \quad (3.37)$$

We obtain (3.33) from the adjugate identity, and at $y = f(x)$ this identity gives $\mathcal{N}_j = \det(G^-) G \delta_j$. Then (3.21), together with (3.26), implies (1). For (2), we multiply the bound on the small numerator by (3.37). Each monomial of the numerator contains at most r matrix coordinates and one vector coordinate. Since P_m contains only matrix coordinates and has degree at most m , this proves (3), including the substitution at the root.

For (4), let $R = (1 + \rho^{-1})/2$, and consider the set

$$\mathcal{G} = \{B^{-1}H : B = B^\top, H = H^\top, I_r \preceq B \preceq C_\psi^2 I_r, p_- B \preceq H \preceq p_+ B\}$$

This set is compact. To see that it contains every population Gram matrix, we divide the known Lebesgue Gram and its density Gram by c_B . It also contains the root parent I_r . Since its members are similar to symmetric matrices with spectrum in $[p_-, p_+]$, the matrix $I_r + 2\rho\zeta t(A) + \rho^2 \zeta^2 I_r$ is invertible on $\mathcal{G} \times \{|\zeta| \leq R\}$.

Recall that the vector moments are uniformly bounded and that $\|\mathcal{T}_j\| \leq 1/2$, including $\mathcal{T}_0 = 0$. By compactness and the continuity of inversion, there exist fixed $\epsilon, M > 0$ such that the $y = 0$, $y = y_0$ and $-\partial_y$ versions of $g_\zeta(G) \det g_\zeta(G^-) \mathcal{N}_j(\theta; y)$ are bounded by M for $|\zeta| \leq R$, $|\theta - \theta_j(x)| \leq \epsilon$ and $|y_0| \leq L_s$, and are holomorphic on a neighborhood of this set. We choose ϵ small enough that the closed neighborhood lies in this holomorphic

domain. Within this neighborhood, perturbations of the matrices may be complex and need not be symmetric.

By Cauchy's formula, each ζ coefficient is bounded by MR^{-l} , where l is its order in the series. Summing the geometric series, we find that all partial sums are bounded by $M' = M/(1 - R^{-1})$, independently of m_j . For nonzero directions, the Cauchy estimate in each direction on $|t_i| = \epsilon/(k|\gamma_i|)$ gives

$$|D^k \Phi(\theta_j)[\gamma_1, \dots, \gamma_k]| \leq M'(k/\epsilon)^k \prod_i |\gamma_i| \leq M' k! (e/\epsilon)^k \prod_i |\gamma_i|.$$

In the last inequality, we used the fact that $k! \geq (k/e)^k$. The cases of zero directions and $k = 0$ are immediate. All compact sets and bounds above depend only on $(r, p_-, p_+, C_\psi, L_s)$, and this proves (4). $\square$

3.4 Unbiased polynomials from a fixed sample

Let $\mathcal{S} = (O_1, \dots, O_{n_b})$ consist of independent observations with common law μ . A polynomial $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^r$ of degree $D \leq n_b$ has a unique homogeneous expansion $\Phi(\theta) = \sum_{k=0}^D \Phi_{(k)}[\theta, \dots, \theta]$, where $\Phi_{(k)}$ is symmetric and k -linear. Using the factorial sums of (3.13), we define its unbiased lift by

$$\widehat{\Phi}(\mathcal{S}; g) = \sum_{k=0}^D \frac{\Sigma_k(\Phi_{(k)}[g(\cdot), \dots, g(\cdot)]; \mathcal{S})}{(n_b)_k}. \quad (3.38)$$

The lift is linear in Φ . If the coefficients of the polynomial are affine in a formal variable y , we lift the coefficients and leave y unchanged. For the indicator of a cell and a monomial of degree k , the lift is the falling factorial of the cell count divided by $(n_b)_k$, the unbiased estimator of a power of a cell probability used in [24, equation (21)].

To compute the variance, we expand around the mean $\theta = \int g d\mu$. This mean enters the analysis, while (3.38) is computed from the known coefficients of the polynomial. For scalar kernels on $\mathcal{O}^k$, let E_i denote integration of the i th argument against μ , and let $\Pi_k = \prod_{i=1}^k (I - E_i)$. Since the operators E_i commute and are orthogonal projections in $L^2(\mu^k)$, we have $\|\Pi_k H\|_2 \leq \|H\|_2$.

Lemma 3.5 (Centered Taylor identity). *Let $g \in L^2(\mu; \mathbb{R}^N)$, $\theta = \int g d\mu$ and $Z(o) = g(o) - \theta$. The lift is square integrable, has mean $\Phi(\theta)$, and satisfies*

$$\widehat{\Phi} - \Phi(\theta) = \sum_{k=1}^D \frac{1}{k!(n_b)_k} \Sigma_k(D^k \Phi(\theta)[Z(\cdot), \dots, Z(\cdot)]; \mathcal{S}). \quad (3.39)$$

For $a \in \mathbb{R}^r$, put

$$\mathcal{K}_k(o_1, \dots, o_k) = a \cdot D^k \Phi(\theta)[g(o_1), \dots, g(o_k)]. \quad (3.40)$$

If $0 < \Lambda \leq n_b - D + 1$, then

$$\text{Var}(a \cdot \widehat{\Phi}) = \sum_{k=1}^D \frac{\|\Pi_k \mathcal{K}_k\|_{L^2(\mu^k)}^2}{k!(n_b)_k} \leq \sum_{k=1}^D \frac{\|\mathcal{K}_k\|_{L^2(\mu^k)}^2}{k!\Lambda^k}. \quad (3.41)$$

For a measurable field of scalar lifts, write $Z(w)$ for the centered lift and $\mathcal{K}_{w,k}$ for its raw kernels, padding with zero up to a common degree D . Let M be a finite measure and suppose these kernels have uniformly bounded $L^2(\mu^k)$ norms. For $v \in L^2(M)$, if a finite partition into blocks B has disjoint observation supports for raw kernels from distinct blocks at every order $k \geq 1$, then

$$\begin{aligned} \left\| \int v(w) Z(w) dM(w) \right\|_{L^2}^2 &\leq \sum_{k=1}^D \frac{\left\| \int v(w) \mathcal{K}_{w,k} dM(w) \right\|_2^2}{k!\Lambda^k} \\ &\leq \max_B M(B) \int v(w)^2 \sum_{k=1}^D \frac{\|\mathcal{K}_{w,k}\|_2^2}{k!\Lambda^k} dM(w). \end{aligned} \quad (3.42)$$

The first inequality also holds without the support assumption.

Proof. By multilinearity and independence, every summand of the lift is integrable, and the lift has the claimed mean. Each summand is square integrable because $g \in L^2$. We expand each homogeneous term using $g = \theta + Z$. For a term of degree l and a choice of k centered arguments, summing over the unused observation indices gives the factor $(n_b - k)_{l-k}$. Since $(n_b)_l = (n_b)_k (n_b - k)_{l-k}$, its coefficient after averaging is $\binom{l}{k} \Phi_{(l)}[Z, \dots, Z, \theta, \dots, \theta]$. Summing over $l \geq k$, we obtain $D^k \Phi(\theta)[Z, \dots, Z]/k!$, which proves (3.39).

Observe that each centered kernel integrates to zero in any one of its arguments. Two summands in (3.39) are therefore orthogonal unless their tuples contain exactly the same observation indices, since an index that appears only once can be integrated first. At order k , there are $(n_b)_k$ choices of the first tuple and $k!$ reorderings for the second. Further, we have $\Pi_k \mathcal{K}_k = a \cdot D^k \Phi(\theta)[Z, \dots, Z]$, and this proves the variance identity. The inequality follows from the fact that Π_k is a contraction and from $(n_b)_k \geq (n_b - D + 1)^k \geq \Lambda^k$.

For the integrated version, since M is finite and the kernel bounds are uniform, we have $\int |v| \|\mathcal{K}_{w,k}\|_2 dM < \infty$. By Minkowski's inequality and Fubini's theorem, we may then integrate the centered Taylor identity and commute the integral with Π_k . We apply the variance identity and the contraction property of the projections to the integrated kernels. At each order, the block integrals are orthogonal because the raw supports are disjoint, and Cauchy-Schwarz on each block gives

$$\left\| \int v \mathcal{K}_{w,k} dM \right\|_2^2 = \sum_B \left\| \int_B v \mathcal{K}_{w,k} dM \right\|_2^2 \leq \sum_B M(B) \int_B v^2 \|\mathcal{K}_{w,k}\|_2^2 dM.$$

Summing over k completes the proof of (3.42). We note that the localization takes place before centering; the actual covariances between different blocks for the fixed sample need not vanish. $\square$

Since each centered kernel integrates to zero in any one of its arguments, the centered Taylor identity is the Hoeffding decomposition [22] of the lift. Our companion paper defines the same lift and proves the same covariance identity [4, Section 4.3, Lemma 4.4]. A variance identity of the same form, for polynomial combinations of complete U-statistics estimating trace moments, appears in [51, Theorem 6.2].

3.5 Definition of the estimator

The estimator uses the independent pilot and evaluation blocks and the comparison parameter Λ of (3.12).

Its parameters are a final level $J \geq 0$, degrees $m_0, \dots, m_J \geq 0$ and a pair level $J_h \geq J$, all of which are integers. We set

$$h = 2^{-J_h}, \quad K_* = K_J = 2^{dJ}, \quad D_j = m_j + q_s \quad (0 \leq j \leq J).$$

We say that the parameters are *admissible* if

$$\begin{aligned} J \geq 0, \quad m_j \geq 0 \quad (0 \leq j \leq J), \quad J_h \geq J, \\ \max_{0 \leq j \leq J} D_j \leq n_b - \Lambda + 1, \quad n_b - 1 \geq \Lambda. \end{aligned} \quad (3.43)$$

The condition on the degrees ensures that the unbiased polynomial estimates below are defined and that their variances satisfy Lemma 3.5. Here $\widehat{V}_n$ denotes the estimator with whichever admissible parameters have been specified. The bounds in Section 3.6 apply to every such choice; the allocation below yields the bound in Theorem 3.1.

Raw features and pilots

For an observation $o = (z, y_o)$, we define the raw features

$$(\mathcal{G}_{j,x}, \beta_{j,x}^f, m_{j,x}^f)(o) = K_j \mathbf{1}_{I_j(x)}(z) (B_j(x)^{-1} \psi_{j,x}(z) \psi_{j,x}(z)^\top, B_j(x)^{-1} \psi_{j,x}(z) y_o, B_j(x)^{-1} \psi_{j,x}(z)). \quad (3.44)$$

Since $\mathbb{E}[Y | X] = f(X)$, their means are $(G_j, \beta_j, m_j)(x)$. For the fixed parent of the root, we use the features $(\mathcal{G}_{-1,x}, \beta_{-1,x}^f, m_{-1,x}^f)(o) = (I_r, 0, 0) \mathbf{1}_{[0,1]^d}(z)$. These have the required means, because $X \in [0,1]^d$ almost surely. We collect the current and parent features in the same order as the moment tuple:

$$g_{j,x} = (\mathcal{G}_{j,x}, \mathcal{G}_{j-1,x}, \beta_{j,x}^f, \beta_{j-1,x}^f, m_{j,x}^f, m_{j-1,x}^f), \quad j \geq 0. \quad (3.45)$$

By (3.8) and Lemma 3.2, these known features are square integrable. Further, $\mathbb{E}g_{j,x} = \theta_j(x)$. Let $g_x = (g_{0,x}, \dots, g_{J,x})$. For each independent pilot block $\mathcal{S} \in \{\mathcal{A}, \mathcal{B}\}$, we apply the lift (3.38) directly to (3.32), keeping y as a formal variable:

$$\widehat{\mathcal{P}}_{x,x'}^{\mathcal{S}}(y) = \widehat{\mathcal{P}}_{x,x'}(\mathcal{S}; g_x)(y) = \sum_{j=0}^J \psi_{j,x}(x')^\top \widehat{F_{j,x}(\cdot; y)}(\mathcal{S}; g_{j,x}). \quad (3.46)$$

The second equality follows from linearity, so only degrees up to $\max_j D_j$ are required. The pilot is affine in y , with coefficients given by

$$u_S = \hat{\mathcal{P}}_{x,x'}^S(0), \quad v_S = -\partial_y \hat{\mathcal{P}}_{x,x'}^S(y), \quad \hat{\mathcal{P}}_{x,x'}^S(y) = u_S - yv_S. \quad (3.47)$$

Pair statistic

The pair statistic uses the kernel k_h of Section 2.2, with $h = 2^{-J_h}$. Since $J_h \geq J$, its pair points satisfy $x' \in I_J(x)$. On the pair set, we define the coefficient vector $c_S = (1, v_S - 1, -u_S)^\top$, so that

$$c_S^\top(y', y, 1)^\top = y' - y - \hat{\mathcal{P}}_{x,x'}^S(y).$$

Let $\hat{N}, \hat{D}$ be given by (2.5), with $m = n_b$ and the evaluation block $\mathcal{P}$. The estimator is

$$\hat{V}_n = \text{clip}_{I_V} \left(\frac{\hat{N}}{\max\{\hat{D}, 1/4\}} \right). \quad (3.48)$$

As in (2.5), a summand is zero off the pair set. The denominator bound in Proposition 2.3 allows us to use the threshold $1/4$, once we verify the pilot hypotheses of that proposition in Proposition 3.6.

Parameters for the refined bound

Recall that $L = \log n$. Approximation at the fine levels requires roughly $\tau m_j \geq \vartheta(T - t_j)$, while the factorial variance costs $\exp\{m_j(t_j + \log m_j)\}$. At the required degree, the leading logarithmic variance cost is largest near $t_j = T/2$. The quadratic profile below keeps the factorial variances within the budget specified below and gives a Gaussian bound for the approximation errors. Park [42, Sections 3.3–3.4] also lowers the degree on finer partitions and balances the geometric approximation error against a factorial variance, and our companion paper uses a related balance [4, Section 4.4]. For $n \geq 3$, we set

$$C_S = 1 + \frac{2(\eta + 2)}{\beta}, \quad S = a_* \sqrt{L} - C_S. \quad (3.49)$$

Since $\beta a_*^2/4 = 2\omega$, we eventually have $S \geq 1$ and

$$\beta S^2/4 + (\eta + 2)S \leq 2\omega L. \quad (3.50)$$

When $S \geq 1$, we define

$$C_D = \max\{1, 4\beta\}, \quad \ell_S = \log(C_0 C_D S), \quad \gamma_* = (r + 1/2)/\vartheta, \quad T^\circ = S - \ell_S - \gamma_* \log S, \quad (3.51)$$

where $C_0 \geq 1$ is the constant fixed in (3.59), and the quadratic profile

$$\mathcal{Q}_S(z) = S/4 + (S - z)^2/S. \quad (3.52)$$

We will need the identities

$$z + \mathcal{Q}_S(z) - S = (z - S/2)^2/S, \quad (3.53)$$

$$z\mathcal{Q}_S(z) = S^2/4 - (S - z)(z - S/2)^2/S \leq S^2/4 \quad (z \leq S). \quad (3.54)$$

The candidate parameters are

$$\begin{aligned} J &= \left\lfloor \frac{L + T^\circ}{d \log 2} \right\rfloor, \quad K_* = 2^{dJ}, \quad T = \log(K_*/n), \\ t_j &= \log(K_j/n), \quad z_j = t_j + \ell_S, \quad u_j = z_j - S/2, \\ D_j &= \lfloor \beta \mathcal{Q}_S(z_j) \rfloor, \quad m_j = D_j - q_S \quad (0 \leq j \leq J), \\ h^\circ &= n^{-2/(d+4)} K_*^{4\vartheta/(d+4)}, \quad J_h = \lceil \log_2(1/h^\circ) \rceil, \quad h = 2^{-J_h}, \\ R_* &= (h^\circ)^2 K_*^{-2\vartheta} = n^{-1} (h^\circ)^{-d/2}. \end{aligned} \quad (3.55)$$

Level-dependent quantities are used only if $J \geq 0$. Rounding gives

$$T^\circ - d \log 2 < T \leq T^\circ, \quad h^\circ/2 < h \leq h^\circ. \quad (3.56)$$

There are $O(L)$ levels, the maximum degree is $O(L^{3/2})$, and the fine levels have degree $O(\sqrt{L})$. The shift ℓ_S accounts for the factorial variance, and the term $\gamma_* \log S$ accounts for the prefactor in the polynomial approximation and for the $O(\sqrt{S})$ contributing levels. Section 3.7 verifies these statements and admissibility.

For every n , the estimator is (3.48) if $n \geq 3$, $S \geq 1$ and (3.43) holds, and v_- otherwise. The estimator defined by this convention is jointly Borel in the observations. Indeed, the known features are Borel, the extraction of polynomial coefficients is a finite algebraic operation, and all factorial sums are finite. The pilots are jointly Borel in their anchors, evaluation points and pilot samples. The pair indicators and the zero convention off pairs are Borel, while the flooring of the denominator and the clipping are continuous. The fallback branch depends only on n and on known constants.

3.6 The pilot and its risk bound

Fix admissible parameters (3.43), and let μ be the law of one observation. For a pair point $w = (x, x')$ with $x' \in I_{J_h}(x)$, we have $|x' - x| \leq \sqrt{d}h$ and

$$hK_*^{1/d} = 2^{J-J_h} \leq 1. \quad (3.57)$$

We denote by M_h the pair measure of (2.3). For either pilot block, we center the coefficient vector:

$$\bar{c} = \mathbb{E}c_{\mathcal{S}}, \quad \zeta_{\mathcal{S}} = c_{\mathcal{S}} - \bar{c}. \quad (3.58)$$

We fix one constant

$$C_0 \geq 1 \quad (3.59)$$

that is sufficiently large for the feature and derivative bounds below, and we define

$$w_j = \sum_{k=1}^{D_j} k! \left(\frac{C_0 K_j}{n} \right)^k, \quad (3.60)$$

$$H_* = \max_{0 \leq j \leq J} w_j \max\{1, n/K_j\}, \quad W = h^2 K_*^{2/d} H_*, \quad (3.61)$$

$$b_h = h \left[K_*^{-\vartheta} + \sum_{j=0}^J K_j^{-\vartheta} (m_j + 1)^r \rho^{m_j} \right]. \quad (3.62)$$

All constants in this subsection depend only on the fixed model parameters, uniformly over levels, anchors and polynomial degrees.

Proposition 3.6 (Polynomial-pilot risk). *For every admissible choice, the pilots and pair statistics are square integrable, $\mathbb{E}_{\theta} \hat{D} \geq 1/2$, and*

$$\sup_{\theta \in \Theta} \|\hat{V}_n - V\|_{L^2} \leq C \{b_h^2 + (n^{-1/2} + n^{-1} h^{-d/2})(1 + W)\}. \quad (3.63)$$

The pilot mean is

$$\mathbb{E} \hat{\mathcal{P}}_{x, x'}^{\mathcal{S}}(y) = \sum_{j=0}^J \psi_{j, x}(x')^{\top} F_j(x; y), \quad (3.64)$$

and its coefficient vector satisfies

$$\sup_w |\bar{c}(w)| \leq C, \quad (3.65)$$

$$\sup_w \mathbb{E} |\zeta_{\mathcal{S}}(w)|^2 \leq CW, \quad (3.66)$$

$$\left\| \int v^{\top} \zeta_{\mathcal{S}} dM_h \right\|_{L^2} \leq C \sqrt{W/n} \|v\|_{L^2(M_h)} \quad (v \in L^2(M_h; \mathbb{R}^3)). \quad (3.67)$$

Proof. The current and parent raw features are supported in $I_j^{\dagger}(x) = I_{\max\{j-1, 0\}}(x)$, and they satisfy

$$|g_{j, x}(o)| \leq CK_j(1 + |y_o|) \mathbf{1}_{I_j^{\dagger}(x)}(z_o). \quad (3.68)$$

We enlarge the fixed constant to cover the root-parent block I_r . By (3.8) and $|I_j^{\dagger}(x)| \leq 2^d / K_j$, we have

$$\int |g_{j, x}|^2 d\mu \leq C_g K_j \quad (3.69)$$

for some fixed constant C_g . The lift lemma implies square integrability and (3.64). On pair points, we have

$$|\psi_{j,x}(x')| \leq ChK_j^{1/d}. \quad (3.70)$$

The order-zero bound in Lemma 3.4 and $\sum_{j=0}^J K_j^{1/d} \leq 2K_*^{1/d}$ imply that both coefficients of the mean pilot are bounded by $ChK_*^{1/d} \leq C$.

For either $\Phi_j = F_{j,x}(\cdot; 0)$ or $\Phi_j = -\partial_y F_{j,x}$, we consider

$$Z_j(w) = \psi_{j,x}(x')^\top \{\hat{\Phi}_j(\mathcal{S}; g_{j,x}) - \Phi_j(\theta_j(x))\}.$$

Its raw kernel of order k is $\psi_{j,x}(x')^\top D^k \Phi_j(\theta_j(x)) [g_{j,x}(o_1), \dots, g_{j,x}(o_k)]$. Using the derivative bound $M_1 k! C_1^k$ and (3.69), we find that the squared norm of this kernel, divided by $k! \Lambda^k$, is at most $Ch^2 K_j^{2/d} k! (6C_1^2 C_g K_j/n)^k$. We choose $C_0 \geq 6C_1^2 C_g$ before fixing the allocation. We partition the pair points by the parent cells of their anchors. Raw kernels from distinct cells have disjoint observation supports, and each block of pair points has M_h -mass at most C/K_j . By (3.42), we have

$$\mathbb{E} Z_j(w)^2 \leq Ch^2 K_j^{2/d} w_j, \quad \left\| \int \nu Z_j dM_h \right\|_2^2 \leq Ch^2 K_j^{2/d-1} w_j \|\nu\|_{L^2(M_h)}^2.$$

Since $w_j \leq H_*$ and $w_j/K_j \leq H_*/n$, Minkowski's inequality and the same geometric sum give (3.66) and (3.67) for the last two coordinates of $\zeta_{\mathcal{S}}$, and its first coordinate is zero. By Tonelli's theorem, we have $\mathbb{E} \int |\zeta_{\mathcal{S}}|^2 dM_h \leq CW < \infty$, which justifies the almost-sure integrals.

For $\mathbf{f} = (f(x'), f(x), 1)^\top$, the mean residual is

$$r_h = \bar{c}^\top \mathbf{f} = f(x') - f(x) - \sum_{j=0}^J \psi_{j,x}(x')^\top F_j(x; f(x)). \quad (3.71)$$

We subtract the telescoping finest fit of Lemma 3.3. Its error is at most $ChK_*^{-\theta}$, and by Lemma 3.4 and (3.70), the remaining sum is at most $Ch \sum_j K_j^{-\theta} (m_j + 1)^r \rho^{m_j}$. We conclude that $\|r_h\|_\infty \leq Cb_h$. These estimates show that the pilot hypotheses of Proposition 2.3 hold, with $b = Cb_h$ and variance budget W . The pilot coefficient fields are jointly Borel by the polynomial construction. Admissibility implies that $n_b \geq 2$ and $n_b \geq n/6$. We apply the proposition with $m = n_b$, $c_* = 1/6$ and $a = 1/4$, and it proves the denominator bound, square integrability and (3.63). Further clipping to I_V fixes V and cannot increase the error. $\square$

3.7 One allocation for bias and variance

We use the quadratic profile and parameters of (3.55). All constants below depend only on (s, d, p_-, p_+, C_0) . By the spatial balance and the dyadic rounding, we have

$$(h^\circ)^{(d+4)/2} = n^{-1} K_*^{2\theta}, \quad R_* = (h^\circ)^2 K_*^{-2\theta} = n^{-1} (h^\circ)^{-d/2}, \quad h^\circ/2 < h \leq h^\circ. \quad (3.72)$$

Since $d\vartheta = s - 1$ and $K_* = ne^T$, we also have

$$(h^\circ)^2 K_*^{2/d} = e^{-2\omega L + \eta T}, \quad R_* = n^{-\lambda} e^{-2(s-1)T/(d+4)}. \quad (3.73)$$

Lemma 3.7 (Allocation). *For all sufficiently large n , the parameters (3.55) are admissible and satisfy*

$$\sum_{j=0}^J K_j^{-\vartheta} (m_j + 1)^r \rho^{m_j} \leq CK_*^{-\vartheta}. \quad (3.74)$$

Consequently,

$$b_h \leq ChK_*^{-\vartheta}, \quad b_h^2 \leq CR_*. \quad (3.75)$$

Moreover, with $c = (d - 4s) / \{4(d + 4)\} > 0$,

$$W \leq 1, \quad n^{-1/2} \leq n^{-c} R_*, \quad (3.76)$$

and

$$n^{-1} h^{-d/2} \leq 2^{d/2} R_*. \quad (3.77)$$

The resulting risk scale is

$$R_* \asymp n^{-\lambda} e^{-\kappa \sqrt{L}} L^{\frac{s-1}{d+4} + \frac{d(q_s-1/2)}{d+4}}. \quad (3.78)$$

Proof. By the choice of S , we have $S \asymp \sqrt{L}$ and, eventually, $\beta S^2/4 + (\eta + 2)S \leq 2\omega L$. The floor in J gives $T = T^\circ + O(1) > 0$ and

$$-L + \ell_S \leq z_j \leq T + \ell_S \leq S - \gamma_* \log S \leq S. \quad (3.79)$$

Since $\mathcal{Q}_S(z) = S/4 + (S - z)^2/S$, all degrees satisfy $D_j \geq \beta S/4 - 1$ and $\max_j D_j = O(L^{3/2}) = o(n)$. Then $J \geq 0$, $m_j \geq 0$, and the sample-size conditions in (3.43) hold eventually. In addition, we have

$$h^2 K_*^{2/d} \leq e^{-2\omega L + \eta T} \leq e^{-\beta S^2/4 - 2S} < 1. \quad (3.80)$$

Since $hK_*^{1/d} = 2^{J-J_h}$, this proves $J_h \geq J$ and hence admissibility.

For the bias, recall that $u_j = z_j - S/2$. Using the identity $z + \mathcal{Q}_S(z) - S = (z - S/2)^2/S$, the degree floor, and $S - \ell_S - T \geq \gamma_* \log S$, we obtain

$$K_j^{-\vartheta} \rho^{m_j} \leq CK_*^{-\vartheta} S^{-r-1/2} e^{-\vartheta u_j^2/S}. \quad (3.81)$$

Here we also used that $\tau\beta = \vartheta$ and $m_j \geq \beta \mathcal{Q}_S(z_j) - q_s - 1$. For $S \geq 1$, we have $\mathcal{Q}_S(z_j) \leq S + 3u_j^2/(2S)$, which implies that $(m_j + 1)^r \leq CS^r(1 + u_j^2/S)^r$. Absorbing this polynomial factor into half of the Gaussian exponent, we find

$$K_j^{-\vartheta} (m_j + 1)^r \rho^{m_j} \leq CK_*^{-\vartheta} S^{-1/2} e^{-\vartheta u_j^2/(2S)}. \quad (3.82)$$

Since the u_j have spacing $d \log 2$, their Gaussian sum is at most $C\sqrt{S}$. This proves (3.74), and (3.75) follows from $h \leq h^\circ$ and (3.72).

For the variance, we abbreviate $x_j = C_0 K_j / n = e^{z_j} / (C_D S)$. When $z \leq 0$ and $S \geq 2$, differentiation shows that $e^z Q_S(z) \leq 5S/4$. This implies that $D_j x_j \leq 5\beta / (4C_D) \leq 5/16 < 1/2$ for $z_j \leq 0$. The factorial series then has ratios of successive terms at most $1/2$, which gives $w_j \leq 2x_j$ and $w_j \max\{1, n/K_j\} \leq 2C_0$. For $0 < z_j \leq S$, the profile gives $D_j \leq 5\beta S/4 \leq C_D S$ and $D_j z_j \leq \beta S^2/4$. Using $k! \leq D_j^k$, we have

$$w_j \leq \sum_{k=1}^{D_j} e^{kz_j} \leq C_D S e^{\beta S^2/4}. \quad (3.83)$$

Since $\max\{1, n/K_j\} \leq C_0 C_D S$ on these fine levels, $H_* \leq 2C_0 C_D^2 S^2 e^{\beta S^2/4}$. The reserve gives

$$W \leq 2C_0 C_D^2 S^2 e^{-2S} \rightarrow 0. \quad (3.84)$$

The pair-noise bound follows from (3.72). Further, since $T \leq S = O(\sqrt{L})$ and $1/2 - \lambda = 2c$, we have $n^{-1/2}/R_* \leq n^{-2c} e^{C\sqrt{L}} \leq n^{-c}$ eventually. This proves (3.76) and (3.77).

Finally, since $\log S = \frac{1}{2} \log L + O(1)$, we have

$$T = a_* \sqrt{L} - \frac{1 + \gamma_*}{2} \log L + O(1).$$

Substituting this into (3.73), and using $2(s-1)a_*/(d+4) = \kappa$ and $\gamma_* = (q_s - 1/2)/\vartheta$, we obtain (3.78). $\square$

3.8 Completion of the upper bound

Proof of Theorem 3.1. The estimator is Borel by the construction in Section 3.5.

For all sufficiently large n , Lemma 3.7 supplies admissible parameters such that $b_h^2 \leq CR_*$, $W \leq 1$, and $n^{-1/2} + n^{-1}h^{-d/2} \leq CR_*$. Inserting these bounds into the master bound (3.63), we obtain

$$\mathfrak{r}_n \leq \sup_{\theta \in \Theta} \|\widehat{V}_n - V\|_{L^2} \leq CR_*.$$

Using the rate identity (3.78), we see that this is exactly the asserted bound, with constants that depend only on the stated model parameters. $\square$

Preliminaries for the lower bounds

In this section, we collect the inputs common to the lower bounds in Theorems 1.1, 1.3 and 1.4: a ternary law, a score-to-risk comparison, and smooth weights whose squares sum to one. We also prove the parametric $n^{-1/2}$ lower bound, which is valid for every $s > 0$. The lower bound in Theorem 1.1 is proved in Section 6, and the lower bound in Theorem 1.3 is proved in Section 5.

4.1 The ternary law

Every construction for the lower bounds uses the following response law, which already appears in the third version of this paper [3, Section 3, equation (3.1)]. Fix constants

$$v_- < v < \min(v_+, \sqrt{C_4}), \quad v < a_Y^2 < C_4/v, \quad a_Y > 0.$$

Such a choice is possible because $v_-^2 < C_4$. For a regression value f and conditional variance V , let Y take values in $\{-a_Y, 0, a_Y\}$ with probabilities

$$q_{\pm a_Y}(f, V) = \frac{V + f^2 \pm a_Y f}{2a_Y^2}, \quad q_0(f, V) = 1 - \frac{V + f^2}{a_Y^2}. \quad (4.1)$$

Then $\mathbb{E}Y = f$, $\text{Var } Y = V$, and

$$\mathbb{E}(Y - f)^4 = a_Y^2 V + (6V - 3a_Y^2)f^2 + 3f^4.$$

At $(f, V) = (0, v)$, all three probabilities are positive and the centered fourth moment is $a_Y^2 v < C_4$. By continuity, there exist $\rho, c > 0$ such that, whenever $|f| \leq \rho$ and $|V - v| \leq \rho$, we have $V \in (v_-, v_+)$, $q_y(f, V) \geq c$ for every response value, and $\mathbb{E}(Y - f)^4 \leq C_4$. This means that the response constraints hold uniformly on this neighborhood. Note also that the probabilities are quadratic in f and affine in V , with $\partial_f^2 q_y = 2\partial_V q_y$. Randomizing the regression mean in order to hide a change of the variance is also the device of the lower-bound proof in [52, Section 4.5], with Gaussian regression values on smooth bumps, and of the fixed-design lower bounds of Wang et al. [56, Section 6.2], who use finite Gaussian moment matching; for the ternary law, the last identity makes this matching exact for one response.

4.2 One score-to-risk comparison

An accurate estimator must follow the target along any path of mixtures. The following lemma bounds how quickly its mean can move. The random variable R_t in the lemma need only represent derivatives of data observables; it need not be the score of the joint law of the latent parameter and the data. Let $J_V = [v_-, \min\{v_+, \sqrt{C_4}\}]$ and $D_V = |J_V|$. The fourth-moment bound implies that every variance in Θ belongs to J_V .

Lemma 4.1 (Score-to-risk comparison). *For $0 \leq t \leq \delta$, let $\mathbb{E}_t$ describe a mixture of n -observation laws with deterministic variance $V_t \in J_V$. Suppose each mixture assigns probability at most ϵ to parameters outside Θ . Suppose also that, for every bounded data-only statistic T , $t \mapsto \mathbb{E}_t T$ is absolutely continuous and, almost everywhere,*

$$\frac{d}{dt} \mathbb{E}_t T = \mathbb{E}_t [T R_t], \quad \mathbb{E}_t R_t = 0.$$

If $B = \int_0^\delta \|R_t\|_{L^2(\mathbb{E}_t)} dt < \infty$ and $\Delta = |V_\delta - V_0|$, then

$$\tau_n^2 \geq \frac{\Delta^2}{(2 + B)^2} - D_V^2 \epsilon. \quad (4.2)$$

Proof. We clip T to J_V , which cannot increase its risk on Θ , and set $r_T^2 = \sup_{\theta \in \Theta} \mathbb{E}_\theta (T - V)^2 + D_V^2 \epsilon$. Since the exceptional states contribute at most $D_V^2 \epsilon$, we have $\mathbb{E}_t (T - V_t)^2 \leq r_T^2$. Then, for $m(t) = \mathbb{E}_t T$, we have

$$|m(t) - V_t| \leq r_T, \quad |m'(t)| = |\mathbb{E}_t [(T - V_t) R_t]| \leq r_T \|R_t\|_{L^2(\mathbb{E}_t)}.$$

Bounding the two endpoint errors and integrating the derivative, we find that $\Delta \leq (2 + B)r_T$. It remains to square this bound and take the infimum over clipped estimators. $\square$

The lemma combines the covariance inequality behind the information bound, which controls the derivative of the mean of a statistic [29, Chapter 2, Theorem 5.10], with a comparison of the means of an estimator under two mixtures, as in the constrained risk inequality for mixtures that Cai and Low [7, Section 2, Theorem 2 and Corollary 1] develop from that of Brown and Low [6]. Here the comparison integrates the derivative along a path of mixtures.

4.3 The parametric lower bound

Proof of the lower bound in Theorem 1.4. We use $p \equiv 1$, $f = 0$, and the ternary law with $V_t = v - a_n t$, $0 \leq t \leq 1$, where $a_n = c_0 n^{-1/2}$. If the fixed constant $c_0 > 0$ is sufficiently

small, then the whole path is admissible for every n . The ordinary likelihood score of this path is

$$R_t = -a_n \sum_{i=1}^n \frac{\partial_V q_{Y_i}(0, V_t)}{q_{Y_i}(0, V_t)}, \quad \mathbb{E}_t R_t^2 = \frac{na_n^2}{V_t(a_Y^2 - V_t)} \leq Cc_0^2.$$

The summands are independent and centered, and the equality follows directly by summing over the three outcomes. By decreasing c_0 if necessary, we can ensure that $B \leq 1$. Applying Lemma 4.1 with $\epsilon = 0$ and $\Delta = a_n$, we obtain

$$\mathfrak{r}_n \geq cn^{-1/2}. \quad (4.3)$$

Since the zero function belongs to every Hölder class, this bound holds for all $s > 0$ and $d \geq 1$. $\square$

4.4 Smooth weights

Both nonparametric lower bounds use one family of smooth weights. As in the third version of this paper [3, Section 1.2, Step 3], their squares sum to one at every point. Let $Q = (-1, 1)^d$, and fix a nonnegative $\psi \in C_c^\infty(Q)$ that is positive on $[-1/2, 1/2]^d$. Then $\sum_{z \in \mathbb{Z}^d} \psi(\cdot - z)^2$ is smooth, 1-periodic and bounded below by a positive constant, and

$$w = \psi \left(\sum_{z \in \mathbb{Z}^d} \psi(\cdot - z)^2 \right)^{-1/2} \in C_c^\infty(Q) \quad \text{satisfies} \quad \sum_{z \in \mathbb{Z}^d} w(u - z)^2 = 1, \quad u \in \mathbb{R}^d. \quad (4.4)$$

We write $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ for the torus and $\|x - x'\|_\infty$ for the sup-distance on it. Let $h = 1/n_h$ with $n_h \geq 4$ an integer, and let $\mathcal{J} = (\mathbb{Z}/n_h\mathbb{Z})^d$. For $j \in \mathcal{J}$, the *patch* Q_j is the set of $x \in \mathbb{T}^d$ with $\|x - hj\|_\infty < h$; for $x \in Q_j$, $(x - hj)/h \in Q$ denotes the rescaled representative of $x - hj$ in $(-h, h)^d$, and the weights are $w_j(x) = w((x - hj)/h)$ for $x \in Q_j$ and $w_j(x) = 0$ otherwise. Because $n_h \geq 4$, distinct points $z \in \mathbb{Z}^d$ with $\|x/h - z\|_\infty < 1$ give distinct $j \in \mathcal{J}$. Then (4.4) implies the partition identity below, and differentiation gives the derivative bounds:

$$\sum_j w_j(x)^2 = 1, \quad \|\partial^\gamma w_j\|_\infty \leq C_\gamma h^{-|\gamma|} \quad \text{for every multi-index } \gamma. \quad (4.5)$$

There are h^{-d} patches, each of volume $(2h)^d$ and meeting at most $3^d - 1$ others. Each point of $\mathbb{T}^d$ lies in at most 2^d patches, and two points of a common patch are at sup-distance less than $2h$.

SECTION 5

The case $0 < s \leq 1$

This section gives a different proof of Theorem 1.3, whose rate was established in the second version of this paper [2, Theorem 1.2 and Appendix A]. Throughout, we assume that $0 < s \leq 1$. All constants may depend on the fixed parameters that define the model class, but not on n .

5.1 Upper bound

The upper bound is the case $t = s$ of Proposition 2.1. The estimator there averages half the squared difference of paired responses whose design points are close, and the proof also covers $s = 1$ and the boundary $d = 4s$. Shen et al. [50, Section 4.1, Proposition 1] prove this rate for their product-kernel ratio estimator when the standardized errors are independent of X ; the proof of Proposition 2.3 allows the error law Q_x to depend on x .

5.2 Lower bound

Proof. If $d \leq 4s$, the parametric bound (4.3) already has the required order. We now suppose that $d > 4s$. We use the smooth weights (4.5) at scale $h = 1/n_h$ with $n_h \geq 4$, and set

$$\eta = bh^s, \quad \xi_j \text{ independent with law } \nu_t = (1-t)\delta_0 + \frac{t}{2}(\delta_{-1} + \delta_1), \quad 0 \leq t \leq 1.$$

Let $p \equiv 1$, $f_\xi = \eta \sum_j w_j \xi_j$ and $V_t = v - \eta^2 t$, and let the responses follow the ternary law. By bounded overlap, we have

$$\|f_\xi\|_\infty \leq C\eta, \quad |f_\xi(x) - f_\xi(x')| \leq C\eta \min\{1, \|x - x'\|/h\} \leq Cb\|x - x'\|^s.$$

For a sufficiently small fixed $b > 0$, every state is then admissible, including at $s = 1$. The periodic weights give the required extension beyond the design cube. We will apply Lemma 4.1 with $\epsilon = 0$ and separation $\Delta = \eta^2$.

Next, we represent derivatives of expectations of bounded statistics of the data along this path, using the latent coefficients. We write $\xi^{j,a}$ for ξ with coordinate j set to a , and define

$$D_j F(\xi) = \frac{1}{2}F(\xi^{j,1}) + \frac{1}{2}F(\xi^{j,-1}) - F(\xi^{j,0}).$$

Since $v'_t = (\delta_{-1} + \delta_1)/2 - \delta_0$, differentiating the finite product prior replaces its j th factor by exactly this signed measure. For a fixed sample, let $L_\xi = \prod_{i=1}^n q_{Y_i}(f_\xi(X_i), V_t)$ and

$$R_t = \sum_j r_j, \quad r_j = \frac{D_j L_\xi - \eta^2 \sum_i w_j(X_i)^2 \partial_{V_i} L_\xi}{L_\xi}, \quad (5.1)$$

where ∂_{V_i} differentiates only the i th factor. Differentiating the prior and the likelihood, and then using $\sum_j w_j^2 = 1$, we obtain $\frac{d}{dt} \mathbb{E}_t T = \mathbb{E}_t [TR_t]$ for all bounded T . Since all sums are finite and the likelihood is uniformly positive for fixed n , the derivative and the integration interchange directly.

Only observations in the support patch Q_j enter r_j , because the other likelihood factors cancel. After this cancellation, summing the remaining numerator over all configurations of the responses in the patch gives zero, since $D_j 1 = 0$ and every response probability sums to one. This implies that $\mathbb{E}_t[r_j \mid \xi, X_1, \dots, X_n] = 0$, and that scores from disjoint patches are conditionally orthogonal. Further, the quadratic ternary probabilities satisfy the exact identity

$$\frac{1}{2} q_y(g + \eta w, V) + \frac{1}{2} q_y(g - \eta w, V) - q_y(g, V) = \eta^2 w^2 \partial_V q_y(g, V).$$

The definition gives $r_j = 0$ when Q_j contains no observation. For one observation $X_i \in Q_j$, this identity, with $w = w_j(X_i)$ and with g excluding the j th field, gives the same conclusion. Indeed, $\partial_V q_y$ does not depend on the mean, so the variance term in (5.1) also equals $\eta^2 w^2 \partial_V q_y(g, V)$, times the likelihood factors of the other observations. This is the cancellation that makes the rate possible. Independent local perturbations of the mean also mask a change of the variance, through finite Gaussian moment matching, in the lower bound of Shen et al. [50, Sections 2.2 and 5], whose design density vanishes between the bumps. Here the ternary law and the identity $\sum_j w_j^2 = 1$ make this cancellation exact for one observation with a design density bounded away from zero.

For the remaining counts, we need only a second-order Taylor estimate, which we include for completeness. Given k observations in Q_j , their response product is $H(a) = \prod_{i=1}^k q_{Y_i}(g_i + \eta w_i a, V_t)$, where g_i excludes the j th field. Along $-1 \leq a \leq 1$, all probabilities are at most one, and their first two derivatives in a are bounded by $C\eta$ and $C\eta^2$, respectively. Then $\sup |H''| \leq Ck^2\eta^2$, which gives $|\{H(1) + H(-1)\}/2 - H(0)| \leq Ck^2\eta^2$. The variance derivative in (5.1) is at most $Ck\eta^2$. Since the denominator is at least c^k , this yields $|r_j|^2 \leq C^k \eta^4$ for $k \geq 2$, after increasing C . For $\mu = nh^d \leq 1$, the binomial count and $|Q_j| \leq Ch^d$ give

$$\mathbb{E}_t r_j^2 \leq \eta^4 \sum_{k \geq 2} \frac{(C\mu)^k}{k!} \leq C\eta^4 \mu^2.$$

There are $O(h^{-d})$ patches, each meeting only a bounded number of others. By conditional orthogonality and $2ab \leq a^2 + b^2$, we have

$$\mathbb{E}_t R_t^2 \leq Ch^{-d} \eta^4 \mu^2 = Cn^2 \eta^4 h^d.$$

We take $h = \lceil 4n^{2/(d+4s)} \rceil^{-1}$. Then $\mu \leq 1$ because $d > 4s$, and $B \leq Cb^2nh^{(d+4s)/2} \leq Cb^2$. We then decrease b so that $B \leq 1$. Applying the score-to-risk comparison, we obtain

$$\mathfrak{r}_n \geq \eta^2/3 \asymp n^{-4s/(d+4s)}.$$

Together with (4.3) and Proposition 2.1, this proves Theorem 1.3, including the boundary $d = 4s$ covered at the outset. $\square$

The lower bound

Throughout this section, we assume that $s > 1$ and that $d > 4s$ is an integer, and the model is that of Section 1.2; in particular, $p_- < 1 < p_+$.

Theorem 6.1. *Let $s > 1$ and let $d > 4s$ be an integer. Let λ, τ and κ be as in (1.2) and (1.7). There are $c > 0$ and n_0 , depending only on $(s, d, p_-, p_+, L_s, v_-, v_+, C_4)$, such that*

$$\mathfrak{r}_n \geq c n^{-\lambda} e^{-\kappa \sqrt{\log n}} (\log n)^{(s-1)/(d+4)} \quad \text{for all } n \geq n_0.$$

The constants below depend only on these fixed model parameters and on the design constants in Section 6.9; they never depend on the sample, the state, n, h, M, D, N or the nuisance values. They also do not depend on $\mathcal{U}$, since the constructed means extend periodically to $\mathbb{R}^d$. We write $L = \log n$.

The ternary law depends on the mean through f and $f^2 + V$. A centered perturbation of the mean imitates an increase of the variance through its diagonal covariance. We correlate the mean with the design density in order to control this covariance at the observed points, and then cancel its diagonal part by decreasing V . Priors under which the covariate density and the regression functions are perturbed together appear in the missing-data lower bound of Robins et al. [47, Section 3]; here the perturbations of the density also select the number of observations in a patch.

The third version of this paper used cardinal interpolation at every count at least two and absorbed signed corrections into a common prior for the design density [3, Section 1.2, Steps 1–2, and Appendix B]. Here we construct the local updates for counts at least three by cardinal interpolation. At count two, the unit and coarse kernels use ordinary selectors, while the fine kernel uses a three-point selector with bounded cell signs. We realize the signed generator through positive weights on update histories, and an n th-power mass tilt cancels the normalization of the likelihood. The risk criterion depends on the activity and the local score energy of the construction, and balancing these two quantities gives the stated rate.

6.1 The sample, patches and raw states

We use the ternary law (4.1) with its constants v, a_Y, ρ . Let c_q denote its uniform positivity constant, and set $\mathcal{Y} = \{-a_Y, 0, a_Y\}$ and $\eta_0 = \rho/2^{d+1}$. Let $P(\mathbf{f}, \mathbf{V}; \mathbf{y}) = \prod_{i \leq k} q_{y_i}(f_i, V_i)$,

with $P = 1$ when $k = 0$. When all variance coordinates equal V , we denote the product by $P(\mathbf{f}, V)$, and the derivative ∂_{V_i} is taken before this substitution. The ternary formulas give

$$\partial_f^2 q_y = 2\partial_V q_y, \quad \sum_{y \in \mathcal{Y}} \partial_V q_y = 0. \quad (6.1)$$

All mark spaces are standard Borel, and densities are Borel versions of their almost-everywhere classes.

We identify $[0, 1]^d$ with $\mathbb{T}^d$ and use the sample space $\mathfrak{X} = (\mathbb{T}^d \times \mathcal{Y})^n$ with the reference measure $\Lambda_n = (dx \otimes \#_{\mathcal{Y}})^{\otimes n}$. For a positive bounded density function p and a mean f , we define

$$U(p, f, V; \xi) = \prod_{i \leq n} p(x_i) q_{y_i}(f(x_i), V), \quad (6.2)$$

$$L(p, f, V; \xi) = m(p)^{-n} U(p, f, V; \xi), \quad m(p) = \int p.$$

When $|f| \leq \rho$ and $|V - v| \leq \rho$, this is the likelihood of n independent observations with design density $p/m(p)$ and the ternary law, and $P_{p,f,V}$ and $\mathbb{E}_{p,f,V}$ denote the corresponding law and expectation. For a patch A , ξ_A consists of the observations that lie in the patch, kept in their original order.

We take the smooth weights (4.5) at $h = 1/n_h$ with $n_h \geq 4$, and set $\mu = nh^d$ and $\omega = \log(1/\mu)$. The chart $\chi_j : Q_j \rightarrow Q = (-1, 1)^d$, $x \mapsto (x - hj)/h$, has Jacobian h^{-d} , and $w_j = w \circ \chi_j$ on Q_j . Recall that each point lies in at most 2^d patches. For distinct labels, we write $j \sim j'$ when $Q_j \cap Q_{j'} \neq \emptyset$. Each label has exactly $3^d - 1$ neighbors, given by the coordinate offsets $\{-1, 0, 1\}^d \setminus \{0\}$ modulo n_h .

For integers $M \geq M_1 = \lceil 4/(p_+ - p_-) \rceil$, let $a_M = p_- + 1/M$ and $b_M = p_+ - 1/M$. The interval between them has width at least $(p_+ - p_-)/2$, and

$$\tau_M = \operatorname{arcosh} \frac{b_M + a_M}{b_M - a_M} = \log \frac{\sqrt{b_M} + \sqrt{a_M}}{\sqrt{b_M} - \sqrt{a_M}}, \quad \tau \leq \tau_M \leq \tau + C_\tau/M. \quad (6.3)$$

Here C_τ is a fixed constant, since differentiating the displayed logarithm gives $\partial_a = \sqrt{b}/[\sqrt{a}(b-a)]$ and $\partial_b = -\sqrt{a}/[\sqrt{b}(b-a)]$, and these derivatives are bounded on the segment from (p_-, p_+) to (a_M, b_M) .

For $D \geq 1$, define

$$I_D = \{\beta \in \mathbb{N}^d : |\beta| \leq D-1\}, \quad \phi_\beta(u) = (u/8)^\beta, \quad \phi = (\phi_\beta)_{\beta \in I_D}, \quad (6.4)$$

where $\mathbb{N} = \{0, 1, \dots\}$ and $\phi_0 = 1$. The fixed constant

$$C_{\text{fr}} = \max\{1, \sup_{\beta} \|w(\cdot)(\cdot/8)^\beta\|_{C^s(\mathbb{R}^d)}\}$$

is finite, because on the compact support of w , the derivatives through order $\ell + 1$ are bounded by $C(1 + |\beta|)^{\ell+1} 8^{-|\beta|}$, and these derivative bounds control the Hölder seminorm.

Let

$$\mathbb{B} = \{c \in \mathbb{R}^{I_D} : \|c\|_1 \leq C_{\text{fr}}^{-1}\}, \quad F_c = \phi^\top c \quad (6.5)$$

Then $|F_c| \leq 1$ and $\|G_c\|_{C^s(\mathbb{R}^d)} \leq 1$ for $G_c = wF_c$, extended by zero outside Q .

A raw state $\vartheta = (p, (c_j)_j) \in \Theta_{\text{raw}}$ consists of a Borel function $p : \mathbb{T}^d \rightarrow [a_M, b_M]$ and coefficients $c_j \in \mathbb{B}$. Its mean is

$$f_\vartheta = \eta_f \sum_j w_j (F_{c_j} \circ \chi_j), \quad \eta_f = c_f h^s, \quad 0 < c_f \leq 1, \quad (6.6)$$

where each term is zero outside Q_j . The vacuum $\vartheta_{\text{vac}}^\zeta$ is the raw state with $p \equiv \zeta$ and $c_j = 0$. The variance is not part of the raw state, and prior probabilities will be placed on marked histories of these states. For $V \in [v - \rho, v]$, we define

$$\begin{aligned} m_\vartheta &= \int p_\vartheta, & \theta(\vartheta, V) &= (p_\vartheta/m_\vartheta, f_\vartheta, V, Q_{\vartheta, V}), \\ Q_{\vartheta, V; x} &= \sum_{y \in \mathcal{Y}} q_y(f_\vartheta(x), V) \delta_{y - f_\vartheta(x)}. \end{aligned} \quad (6.7)$$

We abbreviate the corresponding laws and likelihoods by $P_{\vartheta, V}$, $\mathbb{E}_{\vartheta, V}$, $\mathbb{L}(\vartheta, V; \xi)$ and $\mathbb{U}(\vartheta, V; \xi)$. When $\eta_f \leq \eta_0$, they exist for every raw state, including raw states whose normalized density lies outside the model, and the likelihoods satisfy

$$(a_M c_q / b_M)^n \leq \mathbb{L} \leq (b_M / a_M)^n, \quad (a_M c_q)^n \leq \mathbb{U} \leq b_M^n. \quad (6.8)$$

Let

$$C_w = 2^d \{ \binom{d+\ell}{d} + 2 \}. \quad (6.9)$$

Lemma 6.2 (Legality of raw states). *Let $n_h \geq 4$, $M \geq M_1$, $0 < \eta_f = c_f h^s \leq \eta_0$, $\vartheta \in \Theta_{\text{raw}}$ and $V \in [v - \rho, v]$.*

(a) *The periodic extension satisfies $\|\tilde{f}_\vartheta\|_{C^s(\mathbb{R}^d)} \leq C_w c_f$ and $|f_\vartheta| \leq 2^d \eta_f \leq \rho/2$.*

(b) *The ternary law has mean f_ϑ , variance $V \in (v_-, v_+)$, fourth central moment at most C_4 , and probabilities in $[c_q, 1]$.*

(c) *If $0 < c_m \leq 1/p_+$ and $|m_\vartheta - 1| \leq c_m/M$, then $m_\vartheta > 0$ and $p_- \leq p_\vartheta/m_\vartheta \leq p_+$; if $M \geq 2$, $m_\vartheta \geq 1/2$.*

Consequently, if $C_w c_f \leq L_s$, $M \geq 2$, and the hypothesis of (c) holds, $\theta(\vartheta, V) \in \Theta$.

Proof. The periodic extension is $\tilde{f}_\vartheta(x) = \eta_f \sum_{z \in \mathbb{Z}^d} G_{c_{[z]}}(x/h - z)$. At each point, at most 2^d terms contribute, and for a difference at two points, at most 2^{d+1} terms contribute. Scaling the profile bounds, we obtain

$$\|\partial^\gamma \tilde{f}_\vartheta\|_\infty \leq 2^d c_f \quad (|\gamma| \leq \ell), \quad [D^\ell \tilde{f}_\vartheta]_\alpha \leq 2^{d+1} c_f.$$

Adding the $\binom{d+\ell}{d}$ sup norms proves (a). These bounds also keep the parameters of the response law in the ternary neighborhood, which implies (b). For (c), we have $m_\vartheta \geq 1 - c_m/M > 0$ and

$$\frac{p_- + 1/M}{1 + c_m/M} \geq p_-, \quad \frac{p_+ - 1/M}{1 - c_m/M} \leq p_+.$$

The first inequality uses that $p_- c_m \leq 1$, and the second uses that $p_+ c_m \leq 1$. When $M \geq 2$, we have $m_\theta \geq 1 - 1/M \geq 1/2$. Finally, the normalized density is Borel, the periodic extension restricts to $\mathcal{U}$, and the displayed error kernel is measurable because f_θ is continuous. These facts verify all the model conditions. $\square$

Parameter regime. We set $q_{\text{rsp}} = \max\{2, \lfloor d/(4s) \rfloor\}$ and $C_D = (d+8)/4$. For $K \geq 1$, let $\text{Reg}(K)$ consist of the parameters with

$$\begin{aligned} M \geq M_1 \text{ integer}, \quad D = \lceil C_D M \rceil, \quad N \geq 1, \quad 0 < \mu < 1, \quad \omega = \log(1/\mu), \\ M/K \leq \omega \leq KM, \quad M^2/K \leq \log N \leq KM^2, \\ 0 < \eta_f \leq \eta_0, \quad \eta_f^{4q_{\text{rsp}}} N^d \leq 1. \end{aligned} \quad (6.10)$$

The last inequality bounds the Taylor error of the fixed response rule. Define $C_\sharp = 1/(p_-^2 c_q)$ and $M'_1 = \max\{M_1, \lceil 2/(1-p_-) \rceil, \lceil 2/(p_+ - 1) \rceil\}$. For $M \geq M'_1$, the density interval contains the value one with a fixed margin.

The construction enters the statistical argument through its activity B_{loc} and its local score energy $\mathcal{E}$. For the admissible, annihilating and centered updates constructed below, Proposition 6.15 implies that

$$\tau_n^2 \geq c_{\text{tr}} \frac{\eta_f^4}{1 + B_{\text{loc}}^2 + h^{-d} \mathcal{E}} - (v_+ - v_-)^2 \epsilon_n. \quad (6.11)$$

Here B_{loc} bounds the signed mark weights, $\mathcal{E}$ is defined in (6.70), and ϵ_n is the mass-concentration bound (6.71).

6.2 Interpolation on one patch

In this section, we construct interpolation matrices whose evaluated kernels are diagonal. Their defect and band estimates control the local score energy and the aliases. We work with the frame (6.4) and an integer $d > 4$, and $\|A\|_{\ell^1}$ denotes the sum of the absolute values of the entries of a matrix. A *separated representation* of a permutation-invariant $E : Q^k \rightarrow \mathbb{R}^{I_D \times I_D}$ is

$$E(U) = \int_{\mathcal{Z}} \text{Sym} \left[\prod_{l=1}^k b_l(U_l; \zeta) \right] A(\zeta) \sigma(d\zeta),$$

where $\mathcal{Z}$ is standard Borel, σ is finite and positive, the functions are measurable, $|b_l| \leq 1$, and $\int \|A\|_{\ell^1} d\sigma < \infty$. Here Sym averages over the $k!$ permutations of the points. We call the last integral the *cost* of the representation. If E is symmetric, then symmetrizing A does not increase the cost.

Lemma 6.3. *There are $C_A, C_I \geq 1$, depending only on d , such that for integers $D \geq 2$, $2 \leq k \leq D$ and real $T \geq 1$ there are continuous, permutation-invariant symmetric matrices $E_{k,T}(U)$ and continuous weights $\hat{w}_i(T; U) \in [0, 1]$ satisfying*

$$\phi(U_i)^T E_{k,T}(U) \phi(U_l) = \hat{w}_i(T; U) \mathbf{1}_{\{i=l\}}. \quad (6.12)$$

The weights increase with T . Set $E_{k,T} = \hat{w}_i(T; \cdot) = 0$ for $0 < T < 1$. Each $E_{k,T}$ and each band $E_{k,T} - E_{k,L}$, $0 < L \leq T$, has a separated representation of cost at most

$$C_A^k T^2 (1 + \log T)^{k-2}. \quad (6.13)$$

With $L_+ = \max\{1, L\}$,

$$\int_{Q^k} \|E_{k,T} - E_{k,L}\|_{\ell^1}^2 dU \leq C_I^k L_+^{4-d} (1 + \log L_+)^{k-2}, \quad (6.14)$$

$$\int_{Q^k} \sum_i (1 - \hat{w}_i(T; U)) dU \leq C_A^k T^{-d} (1 + \log T)^{k-2}. \quad (6.15)$$

Proof. For a polynomial in u, u' , let $\|\cdot\|_{\text{coef}}$ be the sum of the absolute values of its coefficients in the monomials $(u/8)^\beta (u'/8)^{\beta'}$. This norm is submultiplicative and equals the ℓ^1 norm of the coefficient matrix. In this proof, all constants C depend only on d . With $\lambda = (8d)^{-1}$ and $\alpha(v) = \lambda|v|^2$, we define

$$f_t(v) = t\alpha(v)^2 e^{-t\alpha(v)}, \quad \int_0^\infty f_t(v) dt = 1 \quad (v \neq 0). \quad (6.16)$$

For $x \neq y$, let $v = y - x$ and $L_{x,y}(u) = v \cdot (y - u) / |v|^2$. The polynomial

$$G_t(x, y)(u, u') = f_t(v) L_{x,y}(u) L_{x,y}(u') = t\lambda^2 e^{-t\lambda|v|^2} [v \cdot (y - u)][v \cdot (y - u')]$$

extends continuously by zero at $x = y$. If Z is a standard Gaussian vector, then differentiating its characteristic function gives

$$G_t(x, y)(u, u') = \frac{\lambda}{2} \sum_{a,b} \mathbb{E} \left[(\delta_{ab} - Z_a Z_b) (y_a - u_a) (y_b - u'_b) \cos(\sqrt{2\lambda t} Z \cdot (y - x)) \right]. \quad (6.17)$$

Expanding the affine factors and the cosine of the difference separates x, y into coordinate monomials times sines or cosines, all of which are bounded by one on Q . The total integrated coefficient norm of these terms is at most C , uniformly in t , because $\sum_{a,b} \mathbb{E} |\delta_{ab} - Z_a Z_b| < \infty$. For these unary representations, the mark spaces are finite unions of Gaussian spaces.

Gaussian integration also gives

$$\sup_x \int_Q f_t(y - x) dy \leq C(1 + t)^{-d/2-1}, \quad \sup_x \int_Q t e^{-t\alpha(y-x)} dy \leq C(1 + t)^{1-d/2}. \quad (6.18)$$

Indeed, since $\alpha(y - x) \leq 1/2$, we have $e^{-t\alpha} \leq e^{1/2}e^{-(1+t)\alpha}$. We then extend the integrals to $\mathbb{R}^d$ and rescale by $\sqrt{\lambda(1+t)}$.

Let $q = k - 1$. For each center i , we assign a scale $t_r \geq 0$ to every $r \neq i$ and write $S = \sum_{r \neq i} \log(1 + t_r)$. We define $E_{k,T}$ through

$$\begin{aligned} \phi(u)^T E_{k,T}(U) \phi(u') &= \sum_i \int_{S \leq 2 \log T} \prod_{r \neq i} G_{t_r}(U_i, U_r)(u, u') dt, \\ \hat{w}_i(T; U) &= \int_{S \leq 2 \log T} \prod_{r \neq i} f_{t_r}(U_r - U_i) dt. \end{aligned} \quad (6.19)$$

The degree is $q \leq D - 1$ in each argument. The coefficients and the weights are continuous because the scale domain is bounded. The cardinal property of $L_{x,y}$ proves (6.12), and a neighbor coincident with the center makes that center's summand zero. The range and the monotonicity of the weights follow from (6.16). Symmetry and permutation invariance are immediate, and $E_{k,1} = 0$.

Next, we multiply the separated representations (6.17), integrate over the scale domain, and average over permutations of the points. Products preserve the bound $|b_l| \leq 1$. Absorbing the coefficient norms into the product Gaussian and scale measures, we obtain a finite positive σ and a measurable A on standard Borel spaces. With $s_r = \log(1 + t_r)$, the cost of these representations is bounded by

$$\frac{kC^q}{(q-1)!} \int_0^{2 \log T} e^S S^{q-1} dS \leq C^k T^2 (1 + \log T)^{k-2}. \quad (6.20)$$

Restricting the scale domain to $2 \log L_+ < S \leq 2 \log T$ gives a representation of the band with the same bound. This proves (6.13).

The remaining estimates use the fact that, for $b \geq 1/2$ and $A \geq 0$,

$$\frac{1}{(q-1)!} \int_A^\infty e^{-bS} S^{q-1} dS = e^{-bA} \sum_{l=0}^{q-1} \frac{A^l}{l! b^{q-l}} \leq C^q e^{-bA} (1 + A)^{q-1}. \quad (6.21)$$

The inequality follows by integration by parts, together with $b^{-1} \leq 2$ and $\sum_l 1/l! \leq e$. Since $\|L_{x,y}\|_{\text{coef}} \leq C/|y - x|$, we have

$$\frac{\|G_t(x, y)\|_{\text{coef}}^2}{f_t(y - x)} \leq C t e^{-t\alpha(y-x)}.$$

Here we define this quotient to be zero where $f_t = G_t = 0$. The measures $dv_i = \prod_{r \neq i} f_{t_r}(U_r - U_i) dt$ have total mass at most one. Applying the Cauchy-Schwarz inequality weighted against their sum, and then (6.18), we obtain, with $A = 2 \log L_+$,

$$\int_{Q^k} \|E_{k,T} - E_{k,L}\|_{\ell^1}^2 dU \leq \frac{C^k}{(q-1)!} \int_A^\infty e^{-(d-4)S/2} S^{q-1} dS.$$

In this bound, we used that the change of variables $s_r = \log(1 + t_r)$ has Jacobian e^S , and the k^2 center factors were absorbed into C^k . Applying (6.21) with $b = (d - 4)/2 \geq 1/2$, we obtain (6.14).

Off the null set of coincident points, $1 - \widehat{w}_i$ is the complementary scale integral. Similarly, the first bound in (6.18) gives

$$\int_{Q^k} \sum_i (1 - \widehat{w}_i) dU \leq \frac{C^k}{(q-1)!} \int_{2 \log T}^{\infty} e^{-dS/2} S^{q-1} dS.$$

Together with (6.21), this implies (6.15). It remains to choose fixed C_A, C_I that cover the preceding bounds. $\square$

6.3 Signed local updates

In this section, we work in one chart, with $u_i \in Q$, $w_i = w(u_i)$, $p_i \in [a_M, b_M]$, $g_i \in [-\rho/2, \rho/2]$, $c \in \mathbb{B}$, $V \in [v - \rho, v]$, $y_i \in \mathcal{Y}$ and $0 < \eta_f \leq \eta_0$. We set

$$f_i(c) = g_i + \eta_f w_i F_c(u_i), \quad P(c) = \prod_{i \leq k} q_{y_i}(f_i(c), V).$$

All local identities below hold uniformly in these nuisance vectors, before they are substituted from a global state in Lemma 6.9. We write $[k] = \{1, \dots, k\}$ and list $\mathbf{u}_S$ in increasing index order. Fix $D \geq M \geq \max\{M_1, 2\}$, and let

$$c_{ab} = (a + b)/2, \quad R_{ab} = (b - a)/2, \quad \tau_{ab} = \text{arcosh}(c_{ab}/R_{ab}),$$

$$\vartheta(z) = \frac{(z - a_M)(b_M - z)}{b_M - a_M}.$$

Note that $0 \leq \vartheta(z) \leq \min\{z - a_M, b_M - z\}$ for $z \in [a_M, b_M]$ and that $\vartheta(0) = -a_M b_M / (b_M - a_M) \neq 0$. The following finite rules are used to select counts and to extract responses. Signed measures that represent polynomial functionals, obtained by duality from polynomial approximation, underlie the moment-matching lower bounds of Lepski, Nemirovski and Spokoiny [30, Section 4.4] and of Wu and Yang [57]; Wu and Yang [58, Appendix A, Lemma 5] apply this duality to the reciprocal on an interval away from zero. Here we select observation counts with the rule in (a), which represents evaluation at the exterior point 0 by a signed measure on the density interval.

Lemma 6.4 (Finite moment rules). (a) For $0 < a < b$ and an integer $n \geq 1$, at $z_i = c_{ab} + R_{ab} \cos(i\pi/n)$, $0 \leq i \leq n$, let $\mathfrak{l}_i = \prod_{j \neq i} z_j / (z_j - z_i)$. Then $\Lambda_n = \sum_i \mathfrak{l}_i \delta_{z_i}$ represents $P \mapsto P(0)$ through degree n , with $\|\Lambda_n\| = \cosh(n\tau_{ab})$. Below $[a, b] = [a_M, b_M]$ and $\tau_{ab} = \tau_M$.

(b) Let n_D be the largest odd integer at most D , and let L_j be the cardinal polynomials at $x_j = \cos(j\pi/n_D)$, $0 \leq j \leq n_D$. The rule $J_D = \sum_j L'_j(0) \delta_{x_j}$ represents $P \mapsto P'(0)$ through degree D , with $\|J_D\| = n_D \leq D$ and symmetric total variation.

(c) At $z_j = j/(2q_{\text{rsp}} + 1)$, $0 \leq j \leq 2q_{\text{rsp}} + 1$, let H_j be the cardinal polynomials. The fixed rule

$$C_* = \sum_j ([z^2] H_j(z)) \delta_{z_j}, \quad \int \varsigma^a dC_* = \mathbf{1}\{a = 2\} \quad (0 \leq a \leq 2q_{\text{rsp}} + 1) \quad (6.22)$$

has finite variation depending only on (s, d) .

(d) For symmetric A , the fixed-atom signed kernel

$$\nu_A = 2C_{\text{fr}}^2 \sum_{\beta, \beta'} A_{\beta\beta'} \mathbb{E}_\iota \left[\iota_1 \iota_2 \delta_{(\iota_1 e_\beta + \iota_2 e_{\beta'}) / (2C_{\text{fr}})} \right], \quad (6.23)$$

where ι_1, ι_2 are independent fair signs, is supported on $\mathbb{B}$ and satisfies

$$\int d\nu_A = 0, \quad \int \mathbf{v} d\nu_A = 0, \quad \int \mathbf{v} \mathbf{v}^\top d\nu_A = A, \quad \|\nu_A\| \leq 2C_{\text{fr}}^2 \|A\|_{\ell^1}. \quad (6.24)$$

Its weights are linear in A .

Proof. Parts (a) and (c) follow from cardinal interpolation. The exterior weights have sign $(-1)^{n+i}$. Applying (a) to the Chebyshev polynomial $T_n((z - c_{ab})/R_{ab})$, whose node values are $(-1)^i$, we find the variation in (a) [45, Section 2.7, Theorem 2.20 and equation (2.37)]. Differentiating the interpolation formula proves (b) through degree n_D . The rule vanishes on even monomials, because reflection pairs nodes with opposite weights. This also gives exactness through degree D when $D = n_D + 1$. With $R(x) = \prod_j (x - x_j)$, the weight $L'_j(0) = -R(0)/(x_j^2 R'(x_j))$ has sign $(-1)^{(n_D-1)/2+j}$, and exactness on T_{n_D} gives the variation in (b). Finally, in (d), averaging over the signs yields zero mass and zero mean. Each summand, before multiplication by its prefactor, has second moment $(e_\beta e_{\beta'}^\top + e_{\beta'} e_\beta^\top)/(4C_{\text{fr}}^2)$. Since A is symmetric, this proves the covariance identity, including repeated indices. Summing the absolute weights proves the bound, and the fixed atoms give measurability. $\square$

For a separated representation $(\mathcal{Z}, \sigma, \mathbf{b}, A)$ of E as in Section 6.2, let $\mathcal{C}(E) = \int \|A\|_{\ell^1} d\sigma$, and define

$$\Gamma_{r,m} = \frac{r^r}{r!} \left(\frac{b_M}{\vartheta(0)} \right)^r \Lambda_{m+r} \otimes J_D^{\otimes r}, \quad \Xi_{r,m,E} = \Gamma_{r,m} \sigma(d\zeta) C_*(d\varsigma) \nu_{A(\zeta)}(d\mathbf{v}). \quad (6.25)$$

For the density mark $e_d = (z, \epsilon, \zeta)$, the update is

$$p_i^e = z + \frac{\vartheta(z) p_i}{r b_M} \sum_{l=1}^r \epsilon_l b_l(u_i; \zeta), \quad c^e = (1 - \varsigma)c + \varsigma \mathbf{v}. \quad (6.26)$$

We keep the density marks separate from the response marks and define

$$\mathcal{B}_A \Phi(c) = \int \Phi((1 - \varsigma)c + \varsigma \mathbf{v}) C_*(d\varsigma) \nu_A(d\mathbf{v}). \quad (6.27)$$

Fix a sufficiently large response constant $C_q \geq 1$, as required below, and let $C_{\text{hi}} = 12p_+ C_q$.

Lemma 6.5 (Local operator realization). *Let $1 \leq r \leq m$ and $r \leq D$, with symmetric A and finite representation cost. The resets are measurable and preserve the density interval and coefficient ball; the density slope has modulus at most $(b_M - a_M)/(4b_M) < 1/4$. Their weights are independent of the incoming state. For every bounded measurable b ,*

$$\int b(e_d) d\Xi_{r,m,E} = 0, \quad \int b(e_d) c^e d\Xi_{r,m,E} = 0. \quad (6.28)$$

Moreover $\|\Gamma_{r,m}\| \leq C^r D^r e^{\tau_M m}$ and $\|\Xi_{r,m,E}\| \leq C^r D^r e^{\tau_M m} \mathcal{C}(E)$.

For every bounded measurable Φ on $\mathbb{B}$ and $k \leq D$,

$$\int \prod_{i \leq k} p_i^e \Phi(c^e) d\Xi_{r,m,E} = \mathfrak{A}_k^{r,m} \sum_{|S|=r} \prod_{i \in S} p_i \mathcal{B}_{E(\mathbf{u}_S)} \Phi(c), \quad (6.29)$$

with zero right side for $k < r$ and

$$\begin{aligned} \mathfrak{A}_k^{r,m} &= \vartheta(0)^{-r} \int z^{k-r} \vartheta(z)^r d\Lambda_{m+r}(z) \quad (k \geq r), \\ \mathfrak{A}_k^{r,m} &= \mathbf{1}\{k = r\} \quad (r \leq k \leq m), \quad |\mathfrak{A}_k^{r,m}| \leq C^k e^{\tau_M m} \quad (k \geq r). \end{aligned} \quad (6.30)$$

For every polynomial Φ of degree at most two, $\mathcal{B}_A \Phi(c) = \frac{1}{2} \text{tr}(AD^2 \Phi(c))$. For the response product $\mathbf{P}$, define

$$\mathcal{Q}_A(c) = \mathcal{B}_A \mathbf{P}(c) - \frac{1}{2} \text{tr}(AD^2 \mathbf{P}(c)).$$

Then

$$|\mathcal{Q}_A| \leq C k^{2q_{\text{rsp}}+2} \eta_f^{2q_{\text{rsp}}+2} \|A\|_{\ell^1}, \quad \mathcal{Q}_A = 0 \quad (k \leq q_{\text{rsp}}). \quad (6.31)$$

If $[\phi(u_i)^T A \phi(u_l)]_{i,l}$ is diagonal with entries d_i ,

$$\mathcal{B}_A \mathbf{P}(c) = \eta_f^2 \sum_i w_i^2 d_i \partial_{V_i} \mathbf{P}(c) + \mathcal{Q}_A(c). \quad (6.32)$$

These response assertions hold for every symmetric A and every count. Uniformly for all $k \geq 1$,

$$|\partial_{V_i} \mathbf{P}| \leq C_q, \quad |\mathcal{B}_A \mathbf{P}(c)| \leq C k^2 \eta_f^2 \|A\|_{\ell^1}. \quad (6.33)$$

More generally, any finite signed update measure Ω with legal resets, density resets measurable in e_d alone, and the conditional annihilation in (6.28) satisfies, for every count,

$$\left| \int \prod_{i \leq k} p_i^e \mathbf{P}(c^e) d\Omega \right| \leq 2C_q k^2 p_+^k \eta_f^2 \|\Omega\|. \quad (6.34)$$

All constants are independent of the frame dimension and counts.

Proof. The averaged modulation has modulus at most one, so the quadratic margin keeps the density admissible, and the slope bound follows from the maximum of the margin. The coefficient reset is a convex combination of two points in the coefficient ball. Conditional annihilation follows from the covariance rule. Multiplying the variations, and using $r^r/r! \leq e^r$ and the fixed rules, we obtain the cost bounds, which also justify every application of Fubini's theorem.

For $k \leq D$, the density product has degree at most D in each ϵ_l . Applying J_D in each coordinate selects r distinct observations and $r!$ assignments, and we obtain

$$\frac{r!}{r^r} \left(\frac{\vartheta(z)}{b_M} \right)^r z^{k-r} \sum_{|S|=r} \prod_{i \in S} p_i \text{Sym} \left[\prod_l b_l(u_{S_l}; \zeta) \right].$$

Then (6.29) follows from the separated representation and the linearity of $\mathcal{B}_A$. The exterior evaluation is exact when $k + r \leq m + r$. For the other counts, the variation bound and the identity $\max \vartheta/|\vartheta(0)| = \sinh^{-2} \tau_M$ give $|\mathfrak{A}_k^{r,m}| \leq b_M^{k-r} \sinh^{-2r}(\tau_M) e^{\tau_M(m+r)}$. Since $M \geq M_1$, (6.3) shows that τ_M is bounded by a fixed constant and is at least τ , which implies (6.30).

For a quadratic Φ , the fixed response rule extracts the quadratic coefficient of $\Phi(c + z(\mathbf{v} - c))$. The zero mass and mean of ν_A leave only $\frac{1}{2} \int D^2 \Phi(c) [\mathbf{v}, \mathbf{v}] d\nu_A(\mathbf{v})$, which is $\frac{1}{2} \text{tr}(AD^2 \Phi(c))$. For the response product, apply the same scalar rule to the Taylor polynomial through degree $2q_{\text{rsp}} + 1$ of $z \mapsto P(c + z(\mathbf{v} - c))$. Its quadratic coefficient, integrated against ν_A , is $\frac{1}{2} \text{tr}(AD^2 P(c))$. The derivative of order $a = 2q_{\text{rsp}} + 2$ along the reset is at most $C_a k^a \eta_f^a$, since each factor has degree two and $|F_v - F_c| \leq 2$. The Taylor remainder and the variation bound for ν_A prove (6.31). Since this scalar product polynomial has degree at most $2k$, the discrepancy vanishes when $k \leq q_{\text{rsp}}$. The heat identity follows from the chain rule and $\partial_{\tilde{f}}^2 q_y = 2\partial_V q_y$.

Finally, let $\Psi(\mathbf{z}) = \prod_i q_{y_i}(g_i + \eta_f w_i z_i, V)$ on $[-1, 1]^k$. Since the ternary derivatives are bounded, we have $|\partial_i \partial_l \Psi| \leq C_q \eta_f^2$ and $|\partial_{V_i} P| \leq C_q$. The remainder of the affine Taylor approximation between any two field vectors is then at most $2C_q k^2 \eta_f^2$. The affine part is annihilated conditionally on the density mark, and this proves both the matrix response bound and (6.34), including counts beyond D . $\square$

6.4 Cancellation for two observations

The unit and coarse kernels use the ordinary selector with $r = 2$. The fine kernel uses a three-point rule whose variation is independent of D , and its higher even modulation terms are controlled by a bounded field in the design density. The regression updates remain smooth.

Lemma 6.6 (A bounded hyperplane field). *For $T \geq 0$ there is a jointly measurable field $\psi_T : Q \rightarrow \{-1, 1\}$, with a law measurable in T , such that $\mathbb{E}\psi_T(u)\psi_T(v) = e^{-T|u-v|}$. Its odd moments vanish. Writing $M_{T,j}(u_1, \dots, u_j) = \mathbb{E} \prod_i \psi_T(u_i)$, for every even $j \geq 4$,*

$$\left\| \mathbb{E} \prod_{i=1}^j \psi_T(u_i) \right\|_{L^2(Q^j)} \leq C^j j^4 T^{-d} \quad (T > 0). \quad (6.35)$$

Proof. Let σ be the uniform probability measure on S^{d-1} , and let $\gamma_d = \int |v_1| d\sigma(v)$. We draw Poisson hyperplanes [27, Section 12.4] with intensity $(T/\gamma_d)\sigma(dv) db$ on $S^{d-1} \times [-\sqrt{d}, \sqrt{d}]$, and then give their cells independent fair signs. There are finitely many hyperplanes. A mark consists of their ordered list and a sign array indexed by the halfspace patterns. With a fixed boundary convention, the evaluation is Borel on a countable union of finite-dimensional spaces. Only the law of the Poisson count depends on T .

Conditionally on the hyperplanes, averaging over the signs leaves one precisely when every occupied cell contains an even number of the indexed points. Two points share a cell when no hyperplane separates them, and the intensity of the hyperplanes that separate them is $(T/\gamma_d) \int |v \cdot (u - u')| d\sigma(v) = T|u - u'|$. This implies the covariance and odd-moment statements. A related representation of the exponential kernel, through the probability that two points lie in a common cell, is derived for STIT tessellations by O'Reilly and Tran [41, Section 4.1, Theorem 4.1]. Such an occupancy partition of $j \geq 4$ points contains two disjoint pairs of indices, with the points of each pair lying in the same cell. For any two chosen pairs, the union of their intervals of separating offsets has length at least half the sum of their lengths. The probability that each of the two pairs lies in a common cell is therefore at most $e^{-T(|u_a - u_b| + |u_c - u_e|)/2}$, and two independent spatial integrations show that the $L^2(Q^j)$ norm of this bound is at most $C^j T^{-d}$. A union bound over the at most j^4 choices of pairs proves (6.35). $\square$

Let $T_0 = Ne^{-c_0 D}$, where $c_0 = \tau + 1$, and suppose that $T_0 > 1$. For $D \geq 3$, we choose fixed symmetric matrices A_0, A_1 with kernels

$$\phi(u)^T A_0 \phi(v) = 1, \quad \phi(u)^T A_1 \phi(v) = -|u - v|^2.$$

Indeed, $A_0 = e_0 e_0^\top$, and $-|u - v|^2 = 2u \cdot v - |u|^2 - |v|^2$. Their coefficient norms depend only on d . Let $\text{Rsp}(A) = C_* \otimes \nu_A$, let $\mathcal{L}_T$ be the law of the marked field, and define

$$X = \frac{1}{2}(\delta_{-1} + \delta_1) - \delta_0, \quad \Gamma_T^{\text{fi}} = \frac{b_M^2}{\vartheta(0)^2} \Lambda_{M+2}(dz) X(d\epsilon) \mathcal{L}_T(d\psi), \quad (6.36)$$

$$p_i^e = z + \frac{\vartheta(z)}{b_M} p_i \epsilon \psi(u_i).$$

Note that $\|X\| = 2$ and that its moments vanish at degree zero and at every odd degree and equal one at every positive even degree. The fine update stays in $[a_M, b_M]$ and has

slope less than $1/4$, by the same margin bound as for (6.26). We use three separate row tags:

$$\begin{aligned}\Xi_{\text{unit}} &= \Gamma_{2,D} \otimes \text{Rsp}(A_0), & \Xi_{\text{co}} &= \int_0^{T_0} T \Gamma_{2,D} \otimes \mathcal{L}_T \otimes \text{Rsp}(A_1) dT, \\ \Xi_{\text{fi}} &= \int_{T_0}^N T \Gamma_T^{\text{fi}} \otimes \text{Rsp}(A_1) dT.\end{aligned}\tag{6.37}$$

For the first two rows, we use the ordinary update (6.26) with $b_1 = b_2 = 1$ and $b_1 = b_2 = \psi$, respectively. Their extraction and admissibility follow from Lemma 6.5. We include T in the mark, so that the displayed integrals are finite signed measures on standard Borel spaces. We denote their direct sum by Ξ_{pair} .

Lemma 6.7 (Pair activity and cancellation). *In $\text{Reg}(K)$, for sufficiently large M , the sum of these rows has variation at most $CN^2e^{\tau_M M}$ and annihilates counts zero and one. At count two, after subtracting the diagonal variance derivative, its numerator is*

$$\eta_f^2 w_1 w_2 p_1 p_2 (1 + Nr) e^{-Nr} \partial_{f_1} \partial_{f_2} P, \quad r = |u_1 - u_2|. \tag{6.38}$$

This expression has squared norm at most $C\eta_f^4 N^{-d}$ for every $N \geq 1$. At $3 \leq k \leq D$ its contribution is $U_k + V_k$, where $U_k = 0$ for $k \leq M$ and

$$\sum_{k=M+1}^D \frac{(C_{\#}\mu)^k}{k!} \|U_k\|_{(k)}^2 \leq C\eta_f^4 e^{2\tau_M M} N^{4-d} \frac{(C\mu)^{M+1}}{(M+1)!}, \tag{6.39}$$

$$\sum_{k=3}^D \frac{(C_{\#}\mu)^k}{k!} \|V_k\|_{(k)}^2 \leq C\eta_f^4 e^{2\tau_M M} \mu^4 T_0^{4-2d} \leq C\eta_f^4 \mu^2 N^{-d}. \tag{6.40}$$

Proof. The variations of the ordinary and fine packets give

$$C\{D^2 e^{\tau_M D} (1 + T_0^2) + N^2 e^{\tau_M M}\} \leq CN^2 e^{\tau_M M} :$$

eventually we have $\tau_M \leq c_0$, so that the coarse term is at most $CD^2 N^2 e^{-c_0 D}$, and the unit term is absorbed because $\log N \asymp M^2$.

For the ordinary pair selector, we write $a_{k,m} = \mathfrak{A}_k^{2,m}$. This coefficient equals one at $k = 2$, vanishes for $3 \leq k \leq m$, and is bounded by $C^k e^{\tau_M m}$. Expanding the fine density product, we obtain the same selected pair term and, for even $4 \leq j \leq k$, the coefficients

$$c_{k,j} = b_M^{2-j} \vartheta(0)^{-2} \int z^{k-j} \vartheta(z)^j d\Lambda_{M+2}(z), \quad |c_{k,j}| \leq C^k e^{\tau_M M}.$$

For the fine row, the vanishing zeroth and first moments of X prove annihilation at counts zero and one. The ordinary rows annihilate these counts by count selection. At count two, the response operator is $p_1 p_2 (\mathcal{B}_{A_0} + W_N(r) \mathcal{B}_{A_1}) P$, where

$$W_N(r) = \int_0^N T e^{-Tr} dT = \frac{1 - (1 + Nr) e^{-Nr}}{r^2}, \quad W_N(0) = N^2/2.$$

At count two, $q_{\text{rsp}} \geq 2$ and (6.31) give exact extraction of the Hessian. The diagonal contribution is the heat derivative, and the cross entry of the evaluated kernel is $1 - r^2 W_N(r) = (1 + Nr)e^{-Nr}$, which proves (6.38). The norm bound follows from $\int_{\mathbb{R}^d} (1 + N\|x\|)^2 e^{-2N\|x\|} dx = CN^{-d}$, and this radial calculation itself holds for every $N \geq 1$.

For higher counts, we define $W^{\text{fi}}(r) = \int_{T_0}^N T e^{-Tr} dT$ and

$$U_k = a_{k,M} \sum_{i < l} p_i p_l W^{\text{fi}}(|u_i - u_l|) \mathcal{B}_{A_1} P.$$

Since $W^{\text{fi}}(r) \leq r^{-2}(1 + T_0 r)e^{-T_0 r}$ and $d > 4$, we have $\|W^{\text{fi}}(|\cdot - \cdot|)\|_2 \leq CT_0^{2-d/2}$. Using the full response bound and the bounded densities, and summing over pairs and responses, we find that $\|U_k\|_{(k)} \leq C^k \eta_f^2 e^{\tau_M M} T_0^{2-d/2}$. The factorial tail starts at $M + 1$. Replacing T_0^{4-d} by N^{4-d} costs a factor $e^{c_0(d-4)D}$, which is absorbed into the fixed C^{M+1} because $D = O(M)$. This proves (6.39).

For even $j \geq 4$, let $H_j = \int_{T_0}^N T M_{T,j} dT$. By Minkowski's inequality and (6.35), we have $\|H_j\|_2 \leq C j^4 T_0^{2-d}$. The remaining numerator is

$$V_k = \sum_{\substack{4 \leq j \leq k \\ j \text{ even}}} c_{k,j} \sum_{|S|=j} \prod_{i \in S} p_i H_j(\mathbf{u}_S) \mathcal{B}_{A_1} P.$$

It follows that $\|V_k\|_{(k)} \leq C^k \eta_f^2 e^{\tau_M M} T_0^{2-d}$, and this term vanishes for $k < 4$. Summing these bounds gives the first inequality in (6.40). Up to a fixed multiplicative constant, the ratio of its right side to $\eta_f^4 \mu^2 N^{-d}$ is $e^{2\tau_M M} \mu^2 N^{4-d} e^{c_0(2d-4)D} = o(1)$, since $\log N \asymp M^2$, $D = O(M)$ and $d > 4$. $\square$

One cutoff for all target counts. Recall that $c_0 = \tau + 1$. For $N \geq 1$ and $\mu \in (0, 1)$, let $H = 1 + \log N$, and define

$$\begin{aligned} \ell_o &= N e^{-c_0 D}, & \nu_r &= N(\mu H)^{(r-2)/d} \quad (2 \leq r \leq D), \\ E_r^{\text{co}} &= E_{r, \min\{\nu_r, \ell_o\}}, & E_r^{\text{fi}} &= E_{r, \nu_r} - E_r^{\text{co}}, & m_r^{\text{co}} &= D, & m_r^{\text{fi}} &= M. \end{aligned} \quad (6.41)$$

As before, $E_{r,T} = 0$ for $T < 1$. We omit the cases $\nu_r < 1$ and the zero bands. For sufficiently large M in $\text{Reg}(K)$, we have $\mu H \leq 1$ and $\ell_o > 1$, and every nonzero fine band has $r \leq M$. The scale calculation in Section 6.6 proves these guard conditions. The coarse guards equal D , so all aliases start at count $M + 1$.

For each selected band with $r \geq 3$ and $b \in \{\text{co}, \text{fi}\}$, we fix a separated representation from Lemma 6.3 with symmetric $A(\zeta)$. For $r = 1$, we take $b = \text{co}$, the constant matrix $E_1 = e_0 e_0^T$ and the guard D . Its representation has one point, with $b_1 \equiv 1$ and $A = e_0 e_0^T$. For each such representation, we form the operator of Lemma 6.5 with its prescribed guard:

$$\Xi_{r,b} = \Xi_{r, m_r^b, E_r^b} \quad (6.42)$$

where $E_1^{\text{co}} = E_1$ and $m_1^{\text{co}} = D$.

At target count two, the row is Ξ_{pair} from (6.37). The singleton, the targets $r \geq 3$ and the three pair components have distinct tags. Let E_0 be the finite disjoint union of their mark spaces, and define

$$\Xi = \Xi_{\text{pair}} \oplus \bigoplus_{r \neq 2, b} \Xi_{r, b}, \quad B_{\text{base}} = \|\Xi\|. \quad (6.43)$$

We also denote the component at target r and band b by Row_r^b . All mark spaces are standard Borel, all updates are measurable and admissible, and the singleton measure is nonzero. Hence $0 < B_{\text{base}} < \infty$.

We write e_d for the density mark, including its row tag. By the conditional annihilation in Lemma 6.5, which also holds for the fine pair row, we have, for every bounded measurable b and $c \in \mathbb{B}$,

$$\int b(e_d) \Xi(de) = 0, \quad \int b(e_d) c^e \Xi(de) = 0. \quad (6.44)$$

A centered reference law for the same measure. The generator needs the signed measure Ξ , but the construction of the prior also needs a positive mark law. We start with $\pi_0 = |\Xi|/B_{\text{base}}$ and the weight $a_0 = B_{\text{base}} d\Xi/d|\Xi|$, where we choose a measurable $\{-1, 1\}$ -valued version of the derivative. Then $\Xi = a_0 \pi_0$ and $|a_0| = B_{\text{base}}$. Note that marks with zero weight may be added without changing any signed integral. We use this freedom to make the reference density have mean one. Let

$$r_{\text{ctr}} = \frac{1}{2} \min\{1 - p_-, p_+ - 1\}, \quad \alpha_{\text{ctr}} = \frac{r_{\text{ctr}}}{2(p_+ - p_-)}, \quad C_{\text{ctr}} = \alpha_{\text{ctr}}^{-1}. \quad (6.45)$$

Lemma 6.8 (The reference marks). *For $M \geq M'_1$, extend E_0 by one zero-weight branch. There is a probability law π on the resulting mark space E and a state-independent weight a° such that*

$$\Xi = a^\circ \pi, \quad B_{\text{loc}} := \sup_E |a^\circ| = C_{\text{ctr}} B_{\text{base}}. \quad (6.46)$$

All updates remain measurable and admissible, with pointwise affine density form $p_i^e = A(e_d, u_i) p_i + B(e_d, u_i)$, $|A| \leq 1$. Their reference means satisfy

$$\int A(e_d, u) \pi(de) = 0, \quad \int B(e_d, u) \pi(de) = 1. \quad (6.47)$$

Thus the expected density after an update is one for every incoming state.

Proof. Under π_0 , the slope has mean zero, since the variation of each ordinary packet factors over symmetric $|J_D|$ modulations. The fine pair packet has the same property because $|X|$ is symmetric. The intercept has a constant mean $\bar{z} \in [a_M, b_M]$. With probability

α_{ctr} , we draw the mark from π_0 and multiply its weight by C_{ctr} . Otherwise, we reset the density to

$$z_0 = \frac{1 - \alpha_{\text{ctr}}\bar{z}}{1 - \alpha_{\text{ctr}}},$$

leave c fixed, and assign weight zero. Since both margins of the interval around one are at least r_{ctr} , the quantities $(1 - \alpha_{\text{ctr}})(z_0 - a_M)$ and $(1 - \alpha_{\text{ctr}})(b_M - z_0)$ are at least $r_{\text{ctr}}/2$, and the reset is admissible. Under π , the slope has mean zero and the intercept has mean one. Only the original branch has nonzero weight, which proves (6.46) and preserves (6.44). All formulas are measurable, and the weight is independent of the state. $\square$

6.5 The generator and its local cancellation

We now transport these concrete updates to the patches. The update $T_{j,e}$ replaces $p_\vartheta(x)$ on Q_j by $A(e_d, \chi_j(x))p_\vartheta(x) + B(e_d, \chi_j(x))$ and c_j by c_j^e , and it leaves everything else unchanged. It preserves raw states. Updates at disjoint patches commute, since neither changes the other's incoming state. For a bounded function G of the raw state, whenever the mark integrands are measurable, we define

$$(\mathcal{A}_j G)(\vartheta) = \int G(T_{j,e}\vartheta) \Xi(de) = \int G(T_{j,e}\vartheta) a^\circ(e) \pi(de), \quad \mathcal{A} = \sum_{j \in \mathcal{J}} \mathcal{A}_j. \quad (6.48)$$

In particular, we have $\|\mathcal{A}_j G\|_\infty \leq B_{\text{loc}} \|G\|_\infty$. For the history notation used later, we also set $a_j(\vartheta, e) = a^\circ(e)$. Every mark integrand used below is measurable by the displayed update formulas, and Section 6.7 makes this explicit for finite histories.

Local score. The operator acts on the raw product U , before normalization. For $\xi = ((x_i, y_i))_{i \leq n}$, we define $\partial_{V_i} U$ by replacing its i th response factor by $\partial_V q_{y_i}$. The *score numerator* and the *local score* are

$$R_j = m_\vartheta^{-n} \left[\mathcal{A}_j U(\cdot, V; \xi)(\vartheta) - \eta_f^2 \sum_{i \leq n} w_j(x_i)^2 \partial_{V_i} U(\vartheta, V; \xi) \right], \quad r_j = R_j / L. \quad (6.49)$$

Here the suppressed arguments are $(\vartheta, V; \xi)$. Since $\sum_j w_j^2 = 1$ and $\sum_i \partial_{V_i} U = \partial_V U$, we have

$$\sum_j R_j = m_\vartheta^{-n} \left[\mathcal{A} U(\cdot, V; \xi)(\vartheta) - \eta_f^2 \partial_V U(\vartheta, V; \xi) \right]. \quad (6.50)$$

We will see that the mass tilt in Section 6.7 identifies the derivative of the mixture with the prior average of this sum.

Local likelihood and energy. For the finite vectors and coefficients fixed in Section 6.3, we define

$$\mathsf{L}_k(\mathbf{p}, c, \mathbf{g}, V; \mathbf{u}, \mathbf{y}) = \prod_{i \leq k} p_i q_{y_i}(f_i(c), V), \quad (6.51)$$

$$\mathsf{R}_k(\mathbf{p}, c, \mathbf{g}, V; \mathbf{u}, \mathbf{y}) = \int_E \mathsf{L}_k(\mathbf{p}^e, c^e, \dots) \Xi(de) - \eta_f^2 \sum_{i \leq k} w_i^2 \partial_{V_i} \mathsf{L}_k(\mathbf{p}, c, \dots), \quad (6.52)$$

where the dots stand for $(\mathbf{g}, V; \mathbf{u}, \mathbf{y})$, $\mathsf{L}_0 = 1$, and $\partial_{V_i} \mathsf{L}_k = \prod_l p_l \cdot (\partial_{V_i} \mathsf{P})(c)$. Since $|f_i(c)| \leq \rho/2 + \eta_0 \leq \rho$, Section 4.1 implies that

$$(a_M c_q)^k \leq \mathsf{L}_k(\mathbf{p}, c, \mathbf{g}, V; \mathbf{u}, \mathbf{y}) \leq b_M^k. \quad (6.53)$$

For a jointly Borel function G that is continuous in the finite-dimensional nuisance variables $(\mathbf{p}, c, \mathbf{g}, V)$, we define

$$\|G\|_{(k)}^2 = \int_{Q^k} \sup_{\mathbf{p}, c, \mathbf{g}, V} \sum_{\mathbf{y} \in \mathcal{Y}^k} |G(\mathbf{p}, c, \mathbf{g}, V; \mathbf{u}, \mathbf{y})|^2 d\mathbf{u}. \quad (6.54)$$

The supremum is taken over the compact ranges stated in Section 6.3. It equals the supremum over a countable dense subset, so the integrand is measurable. This norm applies to all numerators below, since they are polynomials in these nuisance variables with bounded Borel coefficients; the finite variation of Ξ justifies integrating their coefficients. Because the supremum is taken before the integration, the nuisance values may be arbitrary at each tuple $\mathbf{u}$. The triangle inequality for this norm follows from the pointwise supremum and Minkowski's inequality.

Lemma 6.9 (Chart scores and conditional centering). *For the constructed update measure, let $\vartheta \in \Theta_{\text{raw}}$, $V \in [v - \rho, v]$ and $j \in \mathcal{J}$.*

(a) *Suppose ξ_{Q_j} consists of k observations, with chart coordinates $\mathbf{u} \in Q^k$ and responses $\mathbf{y} \in \mathcal{Y}^k$. Relabel these points as $x_i = \chi_j^{-1}(u_i)$ and substitute*

$$p_i = p_{\vartheta}(x_i), \quad c = c_j, \quad g_i = f_{\vartheta}(x_i) - \eta_f w_j(x_i) F_{c_j}(u_i). \quad (6.55)$$

These are admissible nuisance values, and

$$r_j = \frac{\mathsf{R}_k}{\mathsf{L}_k}, \quad |r_j| \leq B_{\text{loc}} \left(\frac{b_M}{a_M c_q} \right)^k + \frac{k \eta_f^2}{a_Y^2 c_q}. \quad (6.56)$$

In particular the score depends on the data only through ξ_{Q_j} .

(b) $\mathsf{R}_0 = 0$ and $\sum_{\mathbf{y}} \mathsf{R}_k(\mathbf{p}, c, \mathbf{g}, V; \mathbf{u}, \mathbf{y}) = 0$ for every k , nuisance vector and design tuple.

(c) *The scores are square integrable and $\mathbb{E}_{\vartheta, V}(r_j \mid X) = 0$, where $X = (x_1, \dots, x_n)$. If $j \neq j'$ are not adjacent, then $\mathbb{E}_{\vartheta, V}(r_j r_{j'} \mid X) = 0$.*

Proof. The background is a sum of other profiles, so $|g_i| \leq (2^d - 1)\eta_f \leq \rho/2$, and the update does not change it. With $I = \{i : x_i \in Q_j\}$, the raw likelihood factors as

$$U = OL_k, \quad O = \prod_{i \notin I} p_\theta(x_i) q_{y_i}(f_\theta(x_i), V). \quad (6.57)$$

Note that the update leaves O fixed and that $w_j = 0$ outside the patch. We conclude that $R_j = m_\theta^{-n} OR_k$ and $L = m_\theta^{-n} OL_k$, which proves the formula for the ratio. The likelihood bounds and $|\partial_V q_y| \leq a_Y^{-2}$ give the envelope in (a). When we sum over responses, the integral term reduces to $\int \prod_i p_i^e d\Xi = 0$ by (6.44), and the variance term also sums to zero. This argument includes $k = 0$ and proves (b). Conditional on X , the response probability is $L_k / \prod_i p_i$, and (b) implies that the score has conditional mean zero. Disjoint patches involve disjoint sets of responses, which are independent given X ; the scores of such patches therefore have zero conditional cross moment. Finally, the envelope in (a), with $k \leq n$, gives square integrability. $\square$

The full response remainder. The full chart numerator also satisfies, for every $k \geq 1$,

$$|R_k(\mathbf{p}, c, \mathbf{g}, V; \mathbf{u}, \mathbf{y})| \leq C_{\text{hi}}^k \eta_f^2 (B_{\text{loc}} + 1). \quad (6.58)$$

Indeed, (6.34) and the bound on the variance derivative give the bound $3C_q k^2 p_+^k \eta_f^2 (B_{\text{loc}} + 1)$, and it remains to use $k^2 \leq 4^k$.

For $2 \leq k \leq D$ and $\mathbf{u} \in Q^k$, we define $\hat{w}_i^{[k]}(\mathbf{u}) = \hat{w}_i(v_k; \mathbf{u})$, with value zero when $v_k < 1$. Recall that $C_\# = 1/(p_-^2 c_q)$.

Proposition 6.10 (Local cancellation and energy). *Let $K \geq 1$. There are $M_0(K)$ and constants C_B, C_5, C_6, C_7 , depending only on $(s, d, p_-, p_+, L_s, v_-, v_+, C_4)$, on the design constants fixed in Section 6.9 and on K , such that for every $(M, D, N, \mu, \eta_f) \in \text{Reg}(K)$ with $M \geq M_0(K)$ the row defined by (6.41) and (6.37) is well defined and has the following properties.*

- (i) *The update measure preserves raw states and has the state-independent representation $\Xi = a^\circ \pi$ constructed above. It obeys (6.44) and the centering identity (6.47).*
- (ii) $B_{\text{loc}} \leq C_B N^2 e^{\tau_M M}$.
- (iii) $R_0 \equiv 0$ and $R_1 \equiv 0$.
- (iv) *For $2 \leq k \leq D$, uniformly in the prescribed nuisance values, $R_k = R_k^{\text{def}} + R_k^{\text{al}} + R_k^{\text{fld}}$. At $k = 2$, the defect is (6.38) and the other terms vanish. For $3 \leq k \leq D$, put*

$$\begin{aligned} R_k^{\text{def}} &= \prod_{l \leq k} p_l \left\{ \mathcal{B}_{E_{k, v_k}(\mathbf{u})} P(c) - \eta_f^2 \sum_{i \leq k} w(u_i)^2 (\partial_{V_i} P)(c) \right\}, \\ R_k^{\text{al}} &= U_k + \sum_{\substack{3 \leq r \leq \min\{k, D\}: E_r^{\text{fi}} \neq 0 \\ M < k}} \mathfrak{A}_k^{r, M} \sum_{S \subset [k], |S|=r} \prod_{l \in S} p_l \mathcal{B}_{E_r^{\text{fi}}(\mathbf{u}_S)} P(c), \\ R_k^{\text{fld}} &= V_k, \end{aligned}$$

where U_k, V_k are the pair alias and field terms in Lemma 6.7. The defect at $k \geq 3$ includes the interpolation error and the fixed response rule's Taylor remainder.

- (v) $\sum_{k=2}^D \frac{(C_\sharp \mu)^k}{k!} \|\mathbf{R}_k^{\text{def}}\|_{(k)}^2 \leq C_5 \eta_f^4 \mu^2 N^{-d}$.
- (vi) $\mathbf{R}_k^{\text{al}} = 0$ for $2 \leq k \leq M$, and $\sum_{k=M+1}^D \frac{(C_\sharp \mu)^k}{k!} \|\mathbf{R}_k^{\text{al}}\|_{(k)}^2 \leq C_6 \eta_f^4 e^{2\tau_M M} N^{4-d} \frac{(C_6 \mu)^{M+1}}{(M+1)!}$.
- (vii) $\sum_{k=3}^D (C_\sharp \mu)^k \|\mathbf{R}_k^{\text{fld}}\|_{(k)}^2 / k! \leq C_7 \eta_f^4 \mu^2 N^{-d}$.

Part (i) follows from the construction and Lemma 6.8. The bounds on the activity and the aliases are proved in Lemmas 6.12 and 6.13, and the field bound is (6.40). For the count decomposition, we note that centering preserves signed likelihood integrals. The score numerators at counts zero and one vanish: only the singleton survives at count one, and its exact heat response cancels the variance derivative. At count two, all other targets vanish, and we are left with (6.38).

The remaining count decomposition follows directly from the ordinary-row identity (6.29). For $3 \leq k \leq D$, the target- k bands sum to E_{k,v_k} , and every other coarse target vanishes. Every other fine target vanishes unless $k > M$; in that case, their targets satisfy $r \leq M < k$, and they give exactly the stated aliases. The pair rows supply $U_k + V_k$. Subtracting the variance derivative proves (iii)–(iv), including the omitted targets where $E_{k,v_k} = 0$. We prove the remaining defect bound next. We emphasize that the activity constant is free of M : an extra factor C^M there would change κ .

Lemma 6.11 (Interpolation and Taylor defects). *Let $N \geq 1$, $\mu \in (0,1)$ and $0 < \eta_f \leq \eta_0$ satisfy $\mu(1 + \log N) \leq 1$ and $\eta_f^{4q_{\text{rsp}}} N^d \leq 1$. Let v_k be as in (6.41), and let $\mathbf{R}_k^{\text{def}}$, $2 \leq k \leq D$, be as in Proposition 6.10(iv). There exists a fixed constant C_5 such that*

$$\sum_{k=2}^D \frac{(C_\sharp \mu)^k}{k!} \|\mathbf{R}_k^{\text{def}}\|_{(k)}^2 \leq C_5 \eta_f^4 \mu^2 N^{-d}.$$

Proof. By the radial integral $\int_{\mathbb{R}^d} (1 + N|x|)^2 e^{-2N|x|} dx = CN^{-d}$, the explicit expression at count two has squared norm at most $C\eta_f^4 N^{-d}$ for every $N \geq 1$. This part needs no additional regime assumption. For $k \geq 3$, the heat identity decomposes the defect into the sum of

$$D_k = -\eta_f^2 \prod_l p_l \sum_i w_i^2 (1 - \hat{w}_i^{[k]}) \partial_{V_i} P, \quad Q_k = \prod_l p_l \mathcal{Q}_{E_{k,v_k}}(\mathbf{u}).$$

If $v_k \geq 1$, interpolation gives

$$\int \sum_i (1 - \hat{w}_i^{[k]}) \leq C^k v_k^{-d} (1 + \log v_k)^{k-2} \leq C^k N^{-d} \mu^{-(k-2)}.$$

If $v_k < 1$, the left side equals $k|Q|^k$, and the same bound follows from $N^d \mu^{k-2} < 1$. Since the defect weights lie in $[0,1]$, we have $(\sum_i (1 - \hat{w}_i^{[k]}))^2 \leq k \sum_i (1 - \hat{w}_i^{[k]})$. Using the

fixed derivative bounds, the nuisance suprema and the sum over responses, we have $\|D_k\|_{(k)}^2 \leq C^k \eta_f^4 N^{-d} \mu^{-(k-2)}$. The corresponding factorial sum is at most $C \eta_f^4 \mu^2 N^{-d}$.

For the Taylor remainder, the band estimate with lower cutoff below one gives $\int \|E_{k,\nu_k}\|_1^2 \leq C^k$, also for omitted targets. Combining this with (6.31), we find $\|Q_k\|_{(k)}^2 \leq C^k k^{4q_{\text{rsp}}+4} \eta_f^{4q_{\text{rsp}}+4}$. The corresponding factorial sum is at most $C \eta_f^{4q_{\text{rsp}}+4} \mu^2$, which is absorbed using $\eta_f^{4q_{\text{rsp}}} N^d \leq 1$. The assertion follows from these two bounds, the pair bound and the squared triangle inequality. $\square$

6.6 Activity and aliases

Fix K . We work in $\text{Reg}(K)$ and take M sufficiently large. In this regime, we have

$$D = O(M), \quad H = O(M^2), \quad M/K \leq \omega \leq KM, \quad \log N \geq M^2/K. \quad (6.59)$$

These bounds imply that $\ell_o > 1$, $\mu H \leq 1$, and $\omega > 2 \log H$ eventually. A nonzero fine band requires $\nu_r > \ell_o$, which implies that

$$r < 2 + \frac{dc_0 D}{\omega - \log H} \leq R,$$

for some fixed integer R depending on K and the model constants. Eventually $R \leq M$, so every fine guard is legal. We choose a sufficiently large fixed C_{act} and define

$$q_a = C_{\text{act}} D H (\mu H)^{a/d}, \quad a = 1, 2. \quad (6.60)$$

Since $\log q_a \leq O(\log M) - aM/(Kd)$, we may choose $M_0(K)$ so that

$$\begin{aligned} M &\geq \max\{M'_1, R\}, & \tau_M &\leq c_0, & \mu H &\leq 1, \\ \log H &\leq M, & q_2 &\leq q_1 \leq \frac{1}{2}, & D^2 q_1 &\leq 1 \end{aligned} \quad (6.61)$$

and all pair bounds hold. Further, any fixed multiple of $D e^{c_0(D+1)}$ is at most N eventually. This proves that the rows are well defined.

Lemma 6.12. *The centered update measure has $B_{\text{loc}} \leq C_B N^2 e^{\tau_M M}$ for a fixed C_B , as asserted in Proposition 6.10(ii).*

Proof. An ordinary target- r band with guard m and representation cost $\mathcal{C}$ has variation at most $C^r D^r e^{\tau_M m} \mathcal{C}$. For $r \geq 3$, using the interpolation cost and $\min\{\nu_r, \ell_o\}^2 \leq \ell_o \nu_r$, we have

$$B_r^{\text{co}} \leq C D^2 e^{\tau_M D} N \ell_o q_1^{r-2}, \quad B_r^{\text{fi}} \leq C D^2 e^{\tau_M M} N^2 q_2^{r-2}.$$

Since $e^{\tau_M D} \ell_o / N \leq 1$, the geometric sums of these bounds are at most $C D^2 N^2 (q_1 + e^{\tau_M M} q_2) \leq C N^2 e^{\tau_M M}$. The singleton costs at most $C D e^{c_0(D+1)} \leq N$. It remains to add the pair bound from Lemma 6.7 and to include the fixed factor from the reference centering. No factor depending exponentially on M is added to the activity. $\square$

Lemma 6.13. *The aliases vanish through count M and, for a fixed C_6 ,*

$$\sum_{k=M+1}^D \frac{(C_\# \mu)^k}{k!} \|\mathbf{R}_k^{\text{al}}\|_{(k)}^2 \leq C_6 \eta_f^4 e^{2\tau_M M} N^{4-d} \frac{(C_6 \mu)^{M+1}}{(M+1)!}.$$

This is Proposition 6.10(vi).

Proof. The pair alias is bounded by (6.39). All other aliases come from nonzero fine bands with $3 \leq r \leq R \leq M$, so none of them occurs for $k \leq M$. Let E be any symmetric matrix kernel. By the uniform response and selector bounds, we have

$$\left\| \mathfrak{A}_k^{r,M} \sum_{|S|=r} \prod_{i \in S} p_i \mathcal{B}_{E(\mathbf{u}_S)} \mathbf{P} \right\|_{(k)} \leq C^k \eta_f^2 e^{\tau_M M} \|E\|_{L^2(\ell^1)}. \quad (6.62)$$

Because the response and selector bounds hold uniformly in the nuisance values at each tuple, they bound the supremum in (6.54). Summing over subsets and responses and integrating over the unused coordinates, we obtain the factor $\binom{k}{r} k^2 3^{k/2} |Q|^{(k-r)/2} \max\{1, p_+\}^k$, which is absorbed in C^k . For a nonzero fine band, interpolation gives

$$\|E_r^{\text{fi}}\|_{L^2(\ell^1)}^2 \leq C^r \ell_\circ^{4-d} (1 + \log \ell_\circ)^{r-2} \leq C^r N^{4-d} e^{CM},$$

where we used $r \leq R$, $D = O(M)$ and $\log H \leq M$. With at most R such targets, their sum satisfies

$$\sum_{k=M+1}^D \frac{(C_\# \mu)^k}{k!} \|\mathbf{R}_k^{\text{al}} - \mathbf{U}_k\|_{(k)}^2 \leq C \eta_f^4 e^{2\tau_M M} N^{4-d} e^{CM} \frac{(C\mu)^{M+1}}{(M+1)!}.$$

Here we used $\sum_{k \geq M+1} x^k/k! \leq e^x x^{M+1}/(M+1)!$ and $\mu < 1$. We absorb e^{CM} into a fixed constant raised to the power $M+1$, and we combine the result with the pair alias by the squared triangle inequality. This changes only the constant in the factorial tail, not the activity exponent. $\square$

6.7 A positive path of priors

In this section, we use the centered marks $(E, \pi, \mathbf{a}^\circ)$ and the updates of Section 6.3, with $M \geq M'_1$, $n_h \geq 4$ and vacuum $\vartheta_{\text{vac}} = \vartheta_{\text{vac}}^1$. Recall that updates at disjoint patches commute, that their weights are state independent, and that $|\mathbf{a}^\circ| \leq B_{\text{loc}}$. We fix the constants

$$\Delta = 3^d, \quad \varrho = 2^{-(\Delta+1)}, \quad c_{\text{pos}} = \frac{1}{2 + 8\Delta^2}, \quad C_{\text{mgf}} = \frac{2^{2d}(p_+ - p_-)^2}{8}. \quad (6.63)$$

We assign polynomial weights to finite marked histories of local updates, and we identify histories that differ only in the order of disjoint updates. The next lemma proves positivity and a differentiation identity. Conditional annihilation then keeps the law of the raw density fixed, and bounded differences control its mass. The interval on which positivity holds depends on B_{loc} , but not on the number of patches.

Lemma 6.14 (Positive realization). *If $\delta B_{\text{loc}}/\varrho \leq c_{\text{pos}}$, there are a standard Borel history space $\mathfrak{H}$, a jointly measurable raw-state map $\vartheta(\mathfrak{h})$ and probability measures ν_t , $0 \leq t \leq \delta$, such that, for every bounded state observable G with measurable history and mark evaluations,*

$$\frac{d}{dt} \int G d\nu_t = \int \mathcal{A}G d\nu_t. \quad (6.64)$$

The product rule holds if G is also continuously differentiable in t in supremum norm. The law of $\mathfrak{m} = \int p_{\vartheta(\mathfrak{h})}$ is independent of t . Moreover

$$\int \mathfrak{m} d\nu_0 = 1, \quad \int e^{u(\mathfrak{m}-1)} d\nu_0 \leq e^{C_{\text{mgf}} u^2 h^d} \quad (u \in \mathbb{R}). \quad (6.65)$$

For fixed $n \geq 1$, define

$$G_n = \int \mathfrak{m}^n d\nu_0, \quad d\sigma_t = G_n^{-1} \mathfrak{m}^n d\nu_t. \quad (6.66)$$

These are probability measures with time-independent mass law, and whenever $\varepsilon \geq 8C_{\text{mgf}} n h^d$,

$$\sigma_t(|\mathfrak{m} - 1| > \varepsilon) \leq 2 \exp \left\{ -\frac{\varepsilon^2}{8C_{\text{mgf}} h^d} \right\}. \quad (6.67)$$

Proof. We call two patch labels dependent if they are equal or adjacent, and independent otherwise. We identify finite words that differ by exchanges of consecutive independent labels, and we call an equivalence class a *shape*. The pieces of a shape are indexed by label and occurrence. We order dependent pieces as in a representative word and take the transitive closure. Shapes are the elements of the free partially commutative monoid of Cartier and Foata [9], and with this order they are heaps of pieces in the sense of Viennot [55, Section 3]. Incomparable pieces have independent labels, and any compatible ordering represents the same shape. To see this, we move the first piece of one ordering to the front of another, passing only incomparable pieces, and repeat. Applying the marked updates therefore gives a well-defined raw state. As the shapes are countable, the disjoint union of their finite products of E is standard Borel. We choose a canonical word for each shape. The finite update recursion along this word makes the density, field, mass and likelihoods jointly measurable, also after a mark is appended.

In the cover graph, two pieces are joined when one covers the other in this partial order. Each piece has at most 2Δ neighbors. Indeed, above and below it there is at most one covering piece of each dependent label, since equal labels are comparable. Maximal

pieces have distinct labels. Removing a maximal piece is the inverse of appending its label and mark, and both operations are measurable finite deletions and insertions of coordinates.

We give each shape g weight $q^{|g|}$ and, conditionally on the shape, we give its pieces independent π marks. This defines a finite reference measure. Indeed, for fixed multiplicities (n_j) , the interleavings of every adjacent pair of label chains determine the shape, and their number is at most

$$\prod_{\{j,k\}: j \sim k} 2^{n_j+n_k} \leq 2^{(\Delta-1) \sum_j n_j}.$$

Summing over the multiplicities, we see that the normalizer is bounded by $\prod_j \sum_{r \geq 0} (q2^{\Delta-1})^r < \infty$, where $q2^{\Delta-1} = 1/4$. Let τ_0 be the normalized reference law, with the empty shape included.

For an upper set U of a history h , let

$$v_t(U) = \text{Leb}\{(s_x)_{x \in U} \in [0, t]^U : s_x < s_y \text{ whenever } x < y\}, \quad v_t(\emptyset) = 1,$$

and define

$$\varphi_t(h) = \sum_{U \text{ upper in } h} q^{-|U|} v_t(U) \prod_{x \in U} a^\circ(e_x). \quad (6.68)$$

Since $v_t(U) = t^{|U|} v_1(U)$, this is a finite measurable polynomial. The weights $v_1(U)$ are the volumes of the order polytopes of the induced posets on U ; see [53, Section 1, Corollary 4.2].

To prove positivity, we let $F(S)$ be the sum in (6.68) restricted to upper sets whose maximal pieces lie in $S \subseteq R$, where $R = \max(h)$. Connected components of an upper set are upper sets, since every upward cover chain stays inside it. Two upper sets intersect exactly when they share a maximal piece of the history. Disjoint upper sets are incomparable, and their order-volume weights multiply. Separating the component P containing $x \in S$, we have

$$F(S) = F(S \setminus \{x\}) + \sum_{\substack{P \text{ connected upper} \\ x \in \max P \subseteq S}} b_t(P) F(S \setminus \max P), \quad |b_t(P)| \leq \epsilon^{|P|}, \quad \epsilon = \frac{tB_{\text{loc}}}{q}.$$

We prove by induction on $|S|$ that $F(S) > 0$ and $1/2 \leq F(S)/F(S \setminus \{x\}) \leq 3/2$. We adapt the deletion induction of Dobrushin [14] for hard-core partition functions, as credited and presented by Scott and Sokal [49, Section 5.1, Theorem 5.1], to the connected upper sets of a history. The induction hypothesis on proper subsets gives $F(S \setminus \max P)/F(S \setminus \{x\}) \leq 2^{|\max P|-1}$. A connected set of k vertices containing x is the vertex set of a depth-first walk of length $2(k-1)$ along a spanning tree, so there are at most $(2\Delta)^{2(k-1)}$ choices. Combining these bounds, we conclude that

$$\left| \frac{F(S)}{F(S \setminus \{x\})} - 1 \right| \leq \sum_{k \geq 1} (8\Delta^2)^{k-1} \epsilon^k = \frac{\epsilon}{1 - 8\Delta^2 \epsilon} \leq \frac{1}{2}.$$

At the stated threshold, the denominator is positive. Multiplying the ratios, we find $2^{-|R|} \leq \varphi_t(\mathfrak{h}) \leq (3/2)^{|R|}$.

Next, we differentiate the order volumes. A coordinate can reach the upper face $s_x = t$ only when x is maximal, and removing it leaves the order volume on the remaining pieces. Because the maximal pieces of an upper set are maximal in the whole history, we have

$$\partial_t \varphi_t(\mathfrak{h}) = \varrho^{-1} \sum_{x \in \max(\mathfrak{h})} a^\circ(e_x) \varphi_t(\mathfrak{h} \setminus x).$$

Appending a specified label adds one independent π mark and multiplies the reference weight of the shape by ϱ . Changing variables between a history and the history obtained by deleting a maximal piece, we obtain

$$\frac{d}{dt} \int G \varphi_t d\tau_0 = \int (\mathcal{A}G) \varphi_t d\tau_0.$$

Here, on histories, $\mathcal{A}$ appends a marked update, and on state observables it is given by (6.48). Since $|R| \leq |\mathcal{J}|$, the bounds on φ_t and its derivative are uniform for fixed n . We then use dominated convergence to justify this identity, its product rule and continuous differentiability in total variation.

Let H be a bounded observable of the raw density. For a fixed incoming state and patch, its updated value depends on the mark only through e_d , and (6.44) then implies that $\mathcal{A}H = 0$. Taking $H = 1$, we find that $dv_t = \varphi_t d\tau_0$ is a probability measure, and general H shows that the law of the raw density is invariant. As $a_M \leq m \leq b_M$, (6.66) then defines probability measures with a constant normalizer and an invariant mass law.

It remains to prove the mass bound under $\tau_0 = \nu_0$. We condition on a shape. Its marks are independent, and by the centering identity (6.47), every fresh density update has expected output one at every point of its patch, for any incoming value. Starting from the vacuum, we find that the conditional mean of the mass is one. We group all marks of each label into one block, so that the blocks are independent. Changing one block changes the final density only on its patch, because all later density updates act pointwise. The resulting change in mass is at most $c = 2^d h^d (p_+ - p_-)$. Revealing these blocks in turn gives centered martingale differences whose conditional ranges have length at most c . The second derivative of the conditional log moment-generating function of each difference is its variance under exponential tilting, which is at most $c^2/4$. Integrating twice, we bound each log moment-generating function by $u^2 c^2/8$. There are h^{-d} blocks, and averaging over the shape proves (6.65). This is the method of bounded differences of McDiarmid [36, Lemma 5.8 and Theorem 6.7], applied conditionally on the shape.

We abbreviate $Z = m - 1$ and $a = C_{\text{mgt}} h^d$. By Jensen's inequality, we have $G_n \geq 1$. Further, $m^n \leq e^{nZ}$. For $\varepsilon \geq 8na$, the exponential Markov inequality with $u = \varepsilon/(2a) - n \geq 0$ gives

$$\sigma_0(Z > \varepsilon) \leq e^{-u\varepsilon} \int e^{(n+u)Z} d\nu_0 \leq e^{n\varepsilon - \varepsilon^2/(4a)} \leq e^{-\varepsilon^2/(8a)}.$$

On the event $Z < -\varepsilon$, we have $m^n \leq 1$, and the same exponential bound gives $\sigma_0(Z < -\varepsilon) \leq e^{-\varepsilon^2/(4a)}$. Adding the two tails and using the invariance of the mass law, we obtain (6.67) for every t . $\square$

6.8 From local updates to risk

Throughout this section, we work with the constructed rows and assume that $M \geq M'_1$, $n_h \geq 4$, $0 < \eta_f = c_f h^s \leq \eta_0$ and $C_w c_f \leq L_s$. We set

$$c_m = \frac{1}{2p_+}, \quad \mathcal{A}_m = \{\mathfrak{h} : |\mathfrak{m}(\mathfrak{h}) - 1| \leq c_m/M\}, \quad (6.69)$$

and define the full local score energy

$$\mathcal{E} = \sum_{k \geq 0} \frac{(C_\# \mu)^k}{k!} \|\mathbf{R}_k\|_{(k)}^2, \quad C_\# = \frac{1}{p_-^2 c_q}, \quad \mu = n h^d. \quad (6.70)$$

Here the numerator and the norm are given by (6.52) and (6.54), and the series is finite by the full-response bound (6.58).

Proposition 6.15 (A lower bound from local cost). *Suppose $c_m/M \geq 8C_{\text{mgt}}\mu$, and put*

$$\epsilon_n = 2 \exp \left\{ -\frac{c_m^2}{8C_{\text{mgt}}M^2 h^d} \right\}. \quad (6.71)$$

Then (6.11) holds with a constant $c_{\text{tr}} > 0$ depending only on the model constants.

Proof. Set

$$I = \sqrt{3^d h^{-d} \mathcal{E}}, \quad C = 1 + B_{\text{loc}} + I, \quad c_* = \min\{1, \varrho c_{\text{pos}}, \rho/\eta_0^2\}, \quad \delta = c_*/C.$$

Then $\delta B_{\text{loc}}/\varrho \leq c_{\text{pos}}$ and $\eta_f^2 \delta \leq \rho$. Lemma 6.14 applies and gives the probability priors (6.66) on $[0, \delta]$. Let $V_t = v - \eta_f^2 t$, and let $\mathbb{E}_t$ denote prior expectation followed by expectation under the normalized observation law at V_t . The choice of the ternary neighborhood and the bound on the variance decrease ensure that $V_t \in J_V$. By Lemma 6.2, every state in $\mathcal{A}_m$ is legal, and the mass tail bound implies that $\sup_{t \leq \delta} \sigma_t(\mathcal{A}_m^c) \leq \epsilon_n$.

For bounded observables T of the data, the mass tilt cancels the likelihood normalization exactly:

$$\mathbb{E}_t T = G_n^{-1} \int \nu_t(d\mathfrak{h}) \int T(\xi) \mathbf{U}(\vartheta(\mathfrak{h}), V_t; \xi) d\Lambda_n(\xi).$$

The inner integral is continuously differentiable in supremum norm for fixed n , since the response factors are affine in V_t and $\Lambda_n(\mathfrak{X}) = 3^n$. Using the product rule in (6.64), we obtain

$$\frac{d}{dt} \mathbb{E}_t T = \mathbb{E}_t [TR], \quad R = \frac{\mathcal{A}\mathbf{U} - \eta_f^2 \partial_V \mathbf{U}}{\mathbf{U}} = \sum_j r_j, \quad \mathbb{E}_t R = 0. \quad (6.72)$$

Here the score identity is (6.50), and the centering follows from Lemma 6.9.

At a fixed raw state, we group the observations according to their count k in Q_j . Using the chart ratio, $dx = h^d du$, and the bound of one on the probability that the remaining observations lie outside the patch, we have

$$\mathbb{E}_{\theta, V} r_j^2 \leq \sum_{k=0}^n \binom{n}{k} \left(\frac{h^d}{m_\theta} \right)^k \int_{Q^k} \sum_{\mathbf{y}} \frac{R_k^2}{L_k} d\mathbf{u} \leq \mathcal{E}.$$

In the second inequality, we used $m_\theta \geq p_-$, $L_k \geq (p_- c_q)^k$, $\binom{n}{k} \leq n^k/k!$ and the uniform bound in (6.54), whose supremum is taken before integration. By Lemma 6.9, disjoint patches have orthogonal full scores. Each label has 3^d dependent labels, including itself, and applying $2ab \leq a^2 + b^2$ to the pairs of dependent labels yields

$$\mathbb{E}_t R^2 \leq 3^d \sum_j \mathbb{E}_t r_j^2 \leq 3^d h^{-d} \mathcal{E} = I^2. \quad (6.73)$$

The integrated score is at most $\delta I \leq c_*$, and the separation is $\eta_f^2 \delta$. Since $D_V \leq v_+ - v_-$, Lemma 4.1 implies that

$$\mathfrak{r}_n^2 \geq \frac{c_*^2 \eta_f^4}{(2 + c_*)^2 \mathcal{C}^2} - (v_+ - v_-)^2 \epsilon_n.$$

Since $\mathcal{C}^2 \leq 3^{d+1}(1 + B_{\text{loc}}^2 + h^{-d} \mathcal{E})$, this proves (6.11) with $c_{\text{tr}} = c_*^2 / [3^{d+1}(2 + c_*)^2]$. $\square$

Lemma 6.16 (Cost of the local construction). *Let $(M, D, N, \mu, \eta_f) \in \text{Reg}(K)$ with $M \geq M_0(K)$, $\mu = nh^d$ and $\eta_f = c_f h^s$. Set $\mathcal{B} = N^2 e^{\tau_M M}$. There is a fixed $C_G \geq 1$, depending only on the model and design constants and K , such that*

$$1 + B_{\text{loc}}^2 + h^{-d} \mathcal{E} \leq C_G \{ \mathcal{B}^2 + \mathfrak{E}_1 + \mathfrak{E}_2 + \mathfrak{E}_3 \}, \quad (6.74)$$

where

$$\begin{aligned} \mathfrak{E}_1 &= n^2 h^{d+4s} N^{-d}, \\ \mathfrak{E}_2 &= nh^{4s} e^{2\tau_M M} N^{4-d} \frac{(C_G \mu)^M}{(M+1)!}, \\ \mathfrak{E}_3 &= nh^{4s} \mathcal{B}^2 \frac{(C_G \mu)^D}{(D+1)!}. \end{aligned} \quad (6.75)$$

Proof. The score numerators at counts zero and one vanish. For $2 \leq k \leq D$, we square the decomposition of Proposition 6.10 into defect, alias and field terms, at the cost of a factor three. For $k > D$, using (6.58), the volume 2^{dk} and the 3^k response vectors, we have

$\|R_k\|_{(k)}^2 \leq (3 \cdot 2^d C_{\text{hi}}^2)^k \eta_f^4 (B_{\text{loc}} + 1)^2$. Let $c_h = 3 \cdot 2^d C_{\text{hi}}^2 C_{\#}$. Using $\sum_{k \geq a} x^k / k! \leq e^x x^a / a!$, we find

$$\begin{aligned} \mathcal{E} &\leq 3(C_5 + C_7) \eta_f^4 \mu^2 N^{-d} \\ &\quad + 3C_6 \eta_f^4 e^{2\tau_M M} N^{4-d} \frac{(C_6 \mu)^{M+1}}{(M+1)!} \\ &\quad + e^{c_h} \eta_f^4 (B_{\text{loc}} + 1)^2 \frac{(c_h \mu)^{D+1}}{(D+1)!}. \end{aligned}$$

We multiply by h^{-d} and use $h^{-d} \mu = n$, $h^{-d} \mu^2 = n^2 h^d$ and $\eta_f = c_f h^s$. As $B_{\text{loc}} \leq C_B \mathcal{B}$ and $\mathcal{B} \geq 1$, the claimed bound holds if we take C_G to be the maximum of

$$1, C_6, c_h, 1 + C_B^2, 3(C_5 + C_7) c_f^4, 3C_6^2 c_f^4, e^{c_h} c_h c_f^4 (C_B + 1)^2.$$

All these constants are fixed before n . □

6.9 Parameters and proof of Theorem 6.1

We choose the parameters by balancing the squared activity with the interpolation contribution $\mathfrak{E}_1$, which leads to $N^{d+4} = n^2 h^{d+4s} e^{-2\tau_M M}$. The remaining contributions are factorial tails. The choice $M \asymp \sqrt{\log n}$ makes these tails smaller than the principal cost, while the correction $-\log M$ in ω produces the logarithmic power in the lower bound. We define

$$\begin{aligned} G &= \frac{d-4s}{d}, & \gamma &= \frac{G}{d+4}, & \theta &= \frac{d\tau}{2(s-1)}, \\ m_*^2 &= \frac{8\gamma(s-1)}{d\tau}, & C_D &= \frac{d+8}{4}, & c_f &= \min\{1, L_s/C_w\}. \end{aligned}$$

Fix K such that θ and γ/m_*^2 lie strictly inside $[1/K, K]$. At this fixed K , we take the constants of Proposition 6.10 and then C_G from Lemma 6.16. Finally, we fix a constant $C_\omega > A + 2$, where

$$A = 1 + \log C_G + \frac{d^2 - 16s}{d(d+4)} \theta + \frac{2d\tau}{d+4}.$$

All these constants are fixed before n . With $L = \log n$, we set

$$\begin{aligned} M &= \lceil m_* \sqrt{L} \rceil, & D &= \lceil C_D M \rceil, & \omega_0 &= \theta M - \log M + C_\omega, & n_h &= \lceil (ne^{\omega_0})^{1/d} \rceil, \\ h &= 1/n_h, & \mu &= nh^d, & \omega &= \log(1/\mu), & \eta_f &= c_f h^s, & N &= \left(n^2 h^{d+4s} e^{-2\tau_M M} \right)^{1/(d+4)}, \\ & & \mathcal{B} &= N^2 e^{\tau_M M}, & R_{\text{low}} &= \eta_f^2 / \mathcal{B}. \end{aligned} \tag{6.76}$$

Lemma 6.17. *For all sufficiently large n , the tuple (M, D, N, μ, η_f) belongs to $\text{Reg}(K)$, and the local construction satisfies every hypothesis of Proposition 6.15. Moreover,*

$$1 + B_{\text{loc}}^2 + h^{-d} \mathcal{E} \leq 3C_G \mathcal{B}^2, \\ R_{\text{low}} \asymp n^{-\lambda} e^{-\kappa \sqrt{L}} L^{(s-1)/(d+4)}, \quad \epsilon_n = o(R_{\text{low}}^2).$$

Proof. Rounding upward gives

$$\omega_0 \leq \omega \leq \omega_0 + d \log 2, \quad \omega = \theta M - \log M + C_\omega + O(1), \quad \tau_M M = \tau M + O(1). \quad (6.77)$$

Since $h^d = e^{-\omega}/n$, we find from the definition of N that

$$(d+4) \log N = GL - (1 + 4s/d)\omega - 2\tau_M M = GL + O(\sqrt{L}). \quad (6.78)$$

We see that $\omega/M \rightarrow \theta$ and $\log N/M^2 \rightarrow \gamma/m_*^2$, which verify the regime bounds at the fixed K . Also, $h \leq n^{-1/d}$ eventually, and since $q_{\text{rsp}} \geq d/(4s) - 1$, we have

$$\eta_f^{4q_{\text{rsp}}} N^d \leq Cn^{(d-4s)/(d+4) - 4q_{\text{rsp}}s/d} \leq Cn^{-4\gamma} \rightarrow 0.$$

All fixed thresholds on M, n_h, η_f hold eventually, and the mass condition $c_m/M \geq 8C_{\text{mgf}}\mu$ follows from $M\mu \rightarrow 0$. The model bound $C_\omega c_f \leq L_s$ holds by construction, and Proposition 6.10 supplies annihilation and reference centering.

For the cost, we have $\mathfrak{E}_1 = \mathcal{B}^2$ exactly. With $c_t = (d^2 - 16s)/\{d(d+4)\}$, direct substitution gives

$$\log \frac{\mathfrak{E}_2}{\mathcal{B}^2} = 4\gamma L + (c_t - M)\omega + \left(\frac{2d\tau_M}{d+4} + \log C_G \right) M - \log(M+1)! \\ = 4\gamma L - M(\omega + \log M) + AM + O(\log M) \leq (-C_\omega + A + o(1))M. \quad (6.79)$$

Here A is independent of C_ω , while the remainder may depend on that subsequently fixed constant. In the final inequality, we used $\theta M^2 \geq 4\gamma L$ and $\omega \geq \omega_0$. This shows that the alias tail tends to zero relative to $\mathcal{B}^2$. For the high-count tail, we have

$$\log \frac{\mathfrak{E}_3}{\mathcal{B}^2} \leq GL - D\omega + O(D) = (G - 4\gamma C_D)L + o(L) = -4\gamma L + o(L).$$

The cost bound follows from Lemma 6.16.

For the risk scale, we compute

$$\log R_{\text{low}} = -\lambda L - \frac{2(s-1)\omega + d\tau_M M}{d+4} + O(1) \\ = -\lambda L - \frac{2d\tau}{d+4} M + \frac{2(s-1)}{d+4} \log M + O(1) \\ = -\lambda L - \kappa \sqrt{L} + \frac{s-1}{d+4} \log L + O(1). \quad (6.80)$$

Finally, $M^2 h^d \leq CL/n$ implies that $\epsilon_n \leq 2e^{-cn/L} = o(R_{\text{low}}^2)$. $\square$

Proof of Theorem 6.1. We take the parameters (6.76) for all sufficiently large n . Combining Lemma 6.17 with Proposition 6.15, we obtain

$$\mathfrak{r}_n^2 \geq \frac{c_{\text{tr}}}{3C_G} R_{\text{low}}^2 - (v_+ - v_-)^2 \epsilon_n \geq \frac{c_{\text{tr}}}{6C_G} R_{\text{low}}^2$$

eventually. Taking square roots and using (6.80) proves the stated lower bound. □

References

- [1] P. M. Aronow and P. Lopatto. On rates attainable under random design: A negative answer to a problem of Robins, 2026. arXiv:2607.13170v1.
- [2] P. M. Aronow and P. Lopatto. On rates attainable under random design: A negative answer to a problem of Robins, 2026. arXiv:2607.13170v2.
- [3] P. M. Aronow and P. Lopatto. Nearly-minimax variance estimation under rough random design, 2026. arXiv:2607.13170v3.
- [4] P. M. Aronow, N. Kallus, and P. Lopatto. Nearly minimax rates for functional estimation under rough random design, 2026. hexagon:2610.00157v1.
- [5] L. D. Brown and M. Levine. Variance estimation in nonparametric regression via the difference sequence method. *The Annals of Statistics*, 35(5):2219–2232, 2007. doi: 10.1214/009053607000000145.
- [6] L. D. Brown and M. G. Low. A constrained risk inequality with applications to nonparametric functional estimation. *The Annals of Statistics*, 24(6):2524–2535, 1996. doi: 10.1214/aos/1032181166.
- [7] T. T. Cai and M. G. Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *The Annals of Statistics*, 39(2): 1012–1041, 2011. doi: 10.1214/10-AOS849.
- [8] T. T. Cai, M. Levine, and L. Wang. Variance function estimation in multivariate nonparametric regression with fixed design. *Journal of Multivariate Analysis*, 100(1): 126–136, 2009. doi: 10.1016/j.jmva.2008.03.007.
- [9] P. Cartier and D. Foata. *Problèmes Combinatoires de Commutation et Réarrangements*, volume 85 of *Lecture Notes in Mathematics*. Springer, Berlin, 1969. doi: 10.1007/BFb0079468.
- [10] X. Chen, L. Liu, and R. Mukherjee. Method-of-moments inference for GLMs and doubly robust functionals under proportional asymptotics, 2025. arXiv:2408.06103v3.
- [11] H. Dette, A. Munk, and T. Wagner. Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(4):751–764, 1998. doi: 10.1111/1467-9868.00152.
- [12] DLMF. NIST Digital Library of Mathematical Functions. <https://dlmf.nist.gov/>, Release 1.2.8 of 2026-09-15, 2026. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B.

- I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
- [13] E. Dobriban, R. Mukherjee, J. M. Robins, and Z. Wang. Improved variance estimation in homoskedastic nonparametric random-design regression via a two-scale approach, 2026. arXiv:2609.08783v1.
- [14] R. L. Dobrushin. Estimates of semi-invariants for the Ising model at low temperatures. In *Topics in Statistical and Theoretical Physics*, volume 177 of *American Mathematical Society Translations, Series 2*, pages 59–81. American Mathematical Society, Providence, RI, 1996. doi: 10.1090/trans2/177/05.
- [15] D. L. Donoho and M. Nussbaum. Minimax quadratic estimation of a quadratic functional. *Journal of Complexity*, 6(3):290–323, 1990. doi: 10.1016/0885-064X(90)90025-9.
- [16] J. Du and A. Schick. A covariate-matched estimator of the error variance in nonparametric regression. *Journal of Nonparametric Statistics*, 21(3):263–285, 2009. doi: 10.1080/10485250802626873.
- [17] J. Fan and Q. Yao. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660, 1998. doi: 10.1093/biomet/85.3.645.
- [18] T. Gasser, L. Sroka, and C. Jennen-Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73(3):625–633, 1986. doi: 10.1093/biomet/73.3.625.
- [19] E. Giné and R. Nickl. A simple adaptive estimator of the integrated square of a density. *Bernoulli*, 14(1):47–61, 2008. doi: 10.3150/07-BEJ110.
- [20] P. Hall and J. S. Marron. On variance estimation in nonparametric regression. *Biometrika*, 77(2):415–419, 1990. doi: 10.1093/biomet/77.2.415.
- [21] P. Hall, J. W. Kay, and D. M. Titterton. Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77(3):521–528, 1990. doi: 10.1093/biomet/77.3.521.
- [22] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948. doi: 10.1214/aoms/1177730196.
- [23] L.-S. Huang and J. Fan. Nonparametric estimation of quadratic regression functionals. *Bernoulli*, 5(5):927–949, 1999. doi: 10.2307/3318450.
- [24] J. Jiao, K. Venkat, Y. Han, and T. Weissman. Minimax estimation of functionals of discrete distributions, 2015. arXiv:1406.6956v5.
- [25] W. Kong and G. Valiant. Estimating learnability in the sublinear data regime. In *Advances in Neural Information Processing Systems*, volume 31, pages 5455–5464, 2018. URL https://papers.nips.cc/paper_files/paper/2018/hash/8bd39eae38511daad6152e84545e504d-Abstract.html. Locators refer to arXiv:1805.01626v3.

- [26] J. K. Kraus, P. S. Vassilevski, and L. T. Zikatanov. Polynomial of best uniform approximation to x^{-1} and smoothing in two-level methods, 2012. arXiv:1002.1859v3.
- [27] G. Last and M. Penrose. *Lectures on the Poisson Process*, volume 7 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge, 2018. doi: 10.1017/9781316104477.
- [28] B. Laurent. Efficient estimation of integral functionals of a density. *The Annals of Statistics*, 24(2):659–681, 1996. doi: 10.1214/aos/1032894458.
- [29] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York, second edition, 1998. doi: 10.1007/b98854.
- [30] O. Lepski, A. Nemirovski, and V. Spokoiny. On estimation of the L_r norm of a regression function. *Probability Theory and Related Fields*, 113(2):221–253, 1999. doi: 10.1007/s004409970006.
- [31] L. Liu, R. Mukherjee, and J. M. Robins. Rejoinder: On nearly assumption-free tests of nominal confidence interval coverage for causal parameters estimated by machine learning. *Statistical Science*, 35(3):545–554, 2020. doi: 10.1214/20-STS804. Section 4 is cited from arXiv:2008.03288v1.
- [32] L. Liu, R. Mukherjee, J. M. Robins, and E. Tchetgen Tchetgen. Adaptive estimation of nonparametric functionals. *Journal of Machine Learning Research*, 22(99):1–66, 2021. URL <https://jmlr.org/papers/v22/liu19-892.html>.
- [33] J. C. Mason and D. C. Handscomb. *Chebyshev Polynomials*. Chapman & Hall/CRC, Boca Raton, FL, 2003.
- [34] R. J. Mathar. Chebyshev series expansion of inverse polynomials. *Journal of Computational and Applied Mathematics*, 196(2):596–607, 2006. doi: 10.1016/j.cam.2005.10.013. Locators refer to arXiv:math/0403344v4.
- [35] A. McClean, S. Balakrishnan, E. H. Kennedy, and L. Wasserman. Double cross-fit doubly robust estimators: Beyond series regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 88(4):1469–1491, 2026. doi: 10.1093/jrsssb/qkag057. Locators refer to arXiv:2403.15175v3.
- [36] C. McDiarmid. On the method of bounded differences. In J. Siemons, editor, *Surveys in Combinatorics*, 1989, pages 148–188. Cambridge University Press, Cambridge, 1989. doi: 10.1017/CBO9781107359949.008.
- [37] R. Mukherjee, W. K. Newey, and J. M. Robins. Semiparametric efficient empirical higher order influence function estimators, 2017. arXiv:1705.07577v1.
- [38] U. U. Müller, A. Schick, and W. Wefelmeyer. Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. *Statistics*, 37(3):179–188, 2003. doi: 10.1080/0233188031000078051.
- [39] A. Munk, N. Bissantz, T. Wagner, and G. Freitag. On difference-based variance estimation in nonparametric regression when the covariate is high dimensional.

- Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):19–41, 2005. doi: 10.1111/j.1467-9868.2005.00486.x.
- [40] W. K. Newey and J. M. Robins. Cross-fitting and fast remainder rates for semiparametric estimation, 2018. arXiv:1801.09138v1.
- [41] E. O’Reilly and N. M. Tran. Stochastic geometry to generalize the Mondrian process. *SIAM Journal on Mathematics of Data Science*, 4(2):531–552, 2022. doi: 10.1137/20M1354490. Locators refer to arXiv:2002.00797v3.
- [42] S. Park. Minimax estimation of the expected conditional covariance under bounds on the covariate density, 2026. arXiv:2610.05006v1.
- [43] J. Rice. Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 12(4):1215–1230, 1984. doi: 10.1214/aos/1176346788.
- [44] T. S. Richardson and A. Rotnitzky. Causal etiology of the research of James M. Robins. *Statistical Science*, 29(4):459–484, 2014. doi: 10.1214/14-STS505.
- [45] T. J. Rivlin. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*. Wiley, New York, second edition, 1990.
- [46] J. Robins, L. Li, E. Tchetgen, and A. van der Vaart. Higher order influence functions and minimax estimation of nonlinear functionals. In D. Nolan and T. Speed, editors, *Probability and Statistics: Essays in Honor of David A. Freedman*, volume 2 of *Institute of Mathematical Statistics Collections*, pages 335–421. Institute of Mathematical Statistics, Beachwood, OH, 2008. doi: 10.1214/193940307000000527.
- [47] J. Robins, E. Tchetgen Tchetgen, L. Li, and A. van der Vaart. Semiparametric minimax rates. *Electronic Journal of Statistics*, 3:1305–1321, 2009. doi: 10.1214/09-EJS479.
- [48] J. M. Robins, L. Li, R. Mukherjee, E. Tchetgen Tchetgen, and A. van der Vaart. Minimax estimation of a functional on a structured high-dimensional model. *The Annals of Statistics*, 45(5):1951–1987, 2017. doi: 10.1214/16-AOS1515.
- [49] A. D. Scott and A. D. Sokal. The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *Journal of Statistical Physics*, 118(5–6):1151–1261, 2005. doi: 10.1007/s10955-004-2055-4. Locators refer to arXiv:cond-mat/0309352v2.
- [50] Y. Shen, C. Gao, D. Witten, and F. Han. Optimal estimation of variance in nonparametric regression with random design. *The Annals of Statistics*, 48(6):3589–3618, 2020. doi: 10.1214/20-AOS1944. Locators refer to arXiv:1902.10822v2.
- [51] X. Song. Batched and complete U-statistics for trace-polynomial estimation from classical shadows, 2026. arXiv:2608.22962v1.
- [52] V. Spokoiny. Variance estimation for high-dimensional regression models. *Journal of Multivariate Analysis*, 82(1):111–133, 2002. doi: 10.1006/jmva.2001.2023. Locators refer to the author manuscript.

- [53] R. P. Stanley. Two poset polytopes. *Discrete & Computational Geometry*, 1(1):9–23, 1986. doi: 10.1007/BF02187680.
- [54] A. van der Vaart. Higher order tangent spaces and influence functions. *Statistical Science*, 29(4):679–686, 2014. doi: 10.1214/14-STS478.
- [55] G. X. Viennot. Heaps of pieces, I: Basic definitions and combinatorial lemmas. In G. Labelle and P. Leroux, editors, *Combinatoire énumérative*, volume 1234 of *Lecture Notes in Mathematics*, pages 321–350. Springer, Berlin, 1986. doi: 10.1007/BFb0072524.
- [56] L. Wang, L. D. Brown, T. T. Cai, and M. Levine. Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, 36(2):646–664, 2008. doi: 10.1214/009053607000000901.
- [57] Y. Wu and P. Yang. Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Transactions on Information Theory*, 62(6):3702–3720, 2016. doi: 10.1109/TIT.2016.2548468.
- [58] Y. Wu and P. Yang. Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *The Annals of Statistics*, 47(2):857–883, 2019. doi: 10.1214/17-AOS1665. Locators refer to arXiv:1504.01227v2.